The relevance of population specific standardisation in the analysis of specific type 2 diabetes mellitus genetic susceptibility loci

BY

GORDON WAYNE TOWERS, M.Sc.

Thesis submitted for the degree Philosophiae Doctor (Ph.D.)
in Biochemistry at the North-West University

PROMOTOR: Professor Antonel Olckers
Centre for Genome Research, North-West University (Potchefstroom Campus)

CO-PROMOTOR: Doctor med. Peter E.H. Schwarz
Department of Endocrinopathies and Metabolic disorders, Technical University Dresden

December 2004

Die toepaslikheid van bevolkingspesifieke standaardisering in die analise van spesifieke tipe 2 diabetes mellitus genetiese vatbaarheidslokusse

DEUR

GORDON WAYNE TOWERS, M.Sc.

Proefskrif voorgelê vir die graad Philosophiae Doctor (Ph.D.)
in Biochemie aan die Noordwes-Universiteit

PROMOTOR: Professor Antonel Olckers
Sentrum vir Genomiese Navorsing, Noordwes-Universiteit (Potchefstroom Kampus)

MEDEPROMOTOR: Doktor med. Peter E.H. Schwarz
Departement Endokrienopatie en Metaboliese afwykings, Tegniese Universiteit Dresden

Desember 2004

This thesis is dedicated to my parents

ABSTRACT

Type 2 diabetes mellitus (T2D) is currently one of the fastest growing non-communicable diseases in the world. It is induced by the pathogenic interaction between insulin resistance and secretion. There are numerous forms of these disorders which are characterised by hyperglycaemia and affect approximately 4% of the general population. This percentage is however rapidly increasing especially in developing regions such as sub-Saharan Africa and Latin America.

During this investigation two diabetic cohorts and two control cohorts consisting of adult black Southern African and Cuban individuals respectively, were screened for reported single nucleotide polymorphisms (SNPs) within the adiponectin and calpain 10 genes. Genotyping was achieved via a real time PCR strategy. Frequency differences between the various genetic configurations of the two cohorts were calculated utilising appropriate statistical analyses.

With regards to the black Southern African cohort, it was determined that certain factors in the calpain 10 gene, e.g. the wild type homozygote at UCSNP-56, were associated with protection towards T2D. Investigation of the Cuban cohort alternatively resulted in the elucidation that this group presents with a differential risk pattern than that of the reported European populations.

Analysis of the adiponectin gene resulted in the determination that within the South African cohort, the G-11391A locus and the 11/12 haplotype combination were associated with protection towards T2D. The variant allele homozygote at the C-11377G locus was associated with increased disease risk within the Cuban cohort investigated. The associations detected in the aforementioned genes were not maintained upon meta-analysis.

When compared to various non-African populations, the investigated SNPs have population specific effects in T2D susceptibility depending on the population investigated. This is most likely due to certain epistatic factors, determination of which will be integral to future investigations of T2D. Data from this investigation indicates that the elucidation and implementation of prevention strategies should be population specific.

OPSOMMING

Tipe 2 diabetes mellitus (T2D) is tans een van die snelgroeienste nie-aansteeklike siektes in die wêreld. Dit word veroorsaak deur die patogeniese wisselwerking tussen insulienweerstand en –vrystelling. Daar is veelvuldige tipes van hierdie siektetoestande wat gekenmerk word deur hiperglisemie en tas ongeveer 4% van die algemene bevolking aan. Hierdie persentasie verhoog egter vinnig, veral in ontwikkelende streke soos sub-Sahara Afrika en Latyns-Amerika.

Tydens hierdie ondersoek is twee diabetiese groepe en twee kontrolegroepe, bestaande uit volwasse swart Suid-Afrikaanse en Kubaanse individue onderskeidelik, gesif vir gerapporteerde enkelnukleotied polimorfismes (SNPs) binne die adiponektien en calpain-10 gene. Genotipering is uitgevoer via 'n kwantitatiewe polimerase kettingreaksie strategie. Frekwensie verskille tussen die genetiese konfigurasies van die twee groepe is bereken deur die gebruik van toepaslike statistiese analises.

Daar is bepaal dat spesifieke faktore in die calpain-10 geen, bv. die homosigoot wat onaangetas is by UCSNP-56, in die swart Suid-Afrikaanse bevolking geassosieer is met beskerming teen T2D. Die ondersoek van die Kubaanse individue het getoon dat 'n ander risikopatroon, as dit wat waargeneem word in die Europese bevolkings, teenwoordig is in hierdie groep.

Analise van die adiponektien geen het gelei tot die gevolgtrekking dat die G-11391A lokus en die 11/12 haplotipe kombinasie geassosieer word met beskerming teen T2D in die swart Suid-Afrikaanse groep. Binne die Kubaanse groep is die homosigoot met 'n aangetaste aleel by die C-11377G lokus geassosieer met verhoogde risiko vir die siekte. Die assosiasies wat waargeneem is in die bogenoemde gene is nie gehandhaaf onder meta-analise toestande nie.

Die ondersoekte SNPs het populasie spesifieke effekte in vatbaarheid vir T2D getoon in vergelyking word met verskillende bevolkings buite Afrika. Dit is moontlik veroorsaak deur sekere epistatiese faktore, waarvan die identifikasie integraal is tot die toekomstige bestudering van T2D. Data van hierdie ondersoek dui daarop dat die ontplooiing en ontwikkeling van voorkomende strategieë bevolkingsspesifiek moet wees.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviations and symbols are presented in alphabetical order.

LIST OF SYMBOLS

1,1	allele combination representing a homozygote for the 1 allele
1,2	allele combination representing a heterozygote
2,2	allele combination representing a homozygote for the 2 allele
°C	degrees Centigrade
>	greater than
<	less than
α	alpha: indicative of significance level when utilised in statistical equations
α'	adjusted significance level
β	beta
χ^2	chi square
δ	delta
ε	epsilon
γ	gamma
κ	kappa: number of independent samples
μ	micro: 10^{-6}
%	percent
Σ	sigma: indicating "the sum of"
τ	tau: measure of inter population variability
T°	temperature
ζ	zeta

LIST OF ABBREVIATIONS

16S rRNA	16S subunit ribosomal ribonucleic acid
2pq	calculated frequency of the 1,2 heterozygous genotype
T45G	alteration of a thymine to a guanine at nucleotide position 45 within the APM1 gene
C-11377G	alteration of a cytosine to a guanine at nucleotide position –11377 within the promoter region of the APM1 gene
G-11391A	alteration of a guanine to an adenine at nucleotide position –11391 within the promoter region of the APM1 gene
A or a	adenine
A_{260}	absorbance of sample at 260 nanometres
A263GG	insertion of two guanines for an adenine at nucleotide position 263 within the HNF1β gene
A3243G	alteration of an adenine to a guanine at nucleotide position 3243 within the $tRNA^{Leu(UUR)}$ gene of the mitochondrial genome
A861G	alteration of an adenine to a guanine at nucleotide position 861 within the leptin receptor gene
ACC	acetyl coenzyme A carboxylase
ACC-β	β-isoform of acetyl coenzyme A carboxylase
ACO	acyl-coenzyme A oxidase
ADA	adenosine deaminase
AFLP	amplified fragment length polymorphism
AIDS	acquired immune deficiency syndrome
Akt	protein kinase B
Ala	alanine
Ala54Thr	alteration of an alanine to a threonine at amino acid residue 54 within the FABP2 protein

List of Abbreviations continued...

Ala931Glu	alteration of an alanine to a glutamic acid at amino acid residue 931 within the PPP1R3A protein
AMPK	5'-adenosine monophosphate activated protein kinase
APM1	adipose most abundant gene transcript 1 or adiponectin
Arg	arginine
Arg127Trp	alteration of arginine to a tryptophan at amino acid residue 127 within the HNF4α protein
Arg154X	alteration of arginine to a stop codon at amino acid residue 154 within the HNF4α protein
Arg197His	alteration of arginine to a histidine at amino acid residue 197 within the HNF4α protein
Arg276X	alteration of arginine to a stop codon at amino acid residue 276 within the HNF1β protein
Asn	asparagine
ASP	agouti signalling protein
Asp	aspartic acid
Asp76Asn	alteration of an aspartic acid to an asparagine at amino acid residue 76 within the IPF-1 protein
Asp276Glu	alteration of an aspartic acid to a glutamic acid at amino acid residue 276 within the HNF4α protein
ATP	adenosine triphosphate
ATPIII	Adult Treatment Panel III
BAC	bacterial artificial chromosome
BMI	body mass index
bp	base pairs
C or c	cytosine
C2792A	alteration of a cytosine to an adenine at nucleotide position 2792 within the PPP1R3A gene
C-terminal	denotes the carboxy terminus of a polypeptide
ca.	circa, approximately
CAD	coronary artery disease
cAMP	cyclic adenosine monophosphate
CAPN10	calpain 10
cDNA	complementary DNA
C/EBP	CCAAT enhancer binding protein
C/EBP B	CCAAT enhancer binding protein β
cGMP	cyclic guanosine monophosphate
CI	confidence interval
c	centi: 10^{-2}
cm	centimetre
cM	centiMorgan
CoA	coenzyme A
COX	cytochrome c oxidase
CPT-1	carnitine palmitoyl transferase-1
Cys	cysteine
Cys18Arg	alteration of a cysteine to an arginine at amino acid residue 18 within the IPF-1 protein
db/db	diabetic/diabetic
df	degrees of freedom
–d(F)\dT	negative differential of the fluorescence level with respect to time
DGGE	denaturing gradient gel electrophoresis
DM	diabetes mellitus
DNA	2'-deoxyribose nucleic acid
dNTP	2'-deoxynucleotide-5'-triphosphate
dTTP	2'-deoxythymidine-5'-triphosphate
dUTP	2'-deoxyuracil-5'-triphosphate
e	base of the natural logarithm = 2.7182818284590452
E	expected population distribution
E. coli	*Escherichia coliform*
Edg	endothelium-cell differentiation family of receptors
EDTA	ethylenediamine tetra-acetic acid: $C_{10}H_{16}N_2O_8$
e.g.	*exempli gratia* meaning "for example"
eIF2B	eukaryotic initiation factor 2B
EM	excitation-maximisation algorithm
F	forward primer
f[1,1]	probability of a 1,1 homozygote
f[1,2]	probability of a heterozygote

List of Abbreviations continued...

f[2,2]	probability of a 2,2 homozygote
F_1	level of differential fluorescence of the peak of the 1,1 homozygote
F_2	level of differential fluorescence of the peak of the 2,2 homozygote
F_3	level of differential fluorescence of the right hand peak of the 1,2 heterozygote
F_4	level of differential fluorescence of the left hand peak of the 1,2 heterozygote
FA	fatty acids
FABP2	fatty acid binding protein 2
FAS	fatty acid synthetase
FFA	free fatty acid
FISH	fluorescent *in situ* hybridisation
FPG	fasting plasma glucose
FRET	fluorescence resonance energy transfer
G or g	guanine
g	gram
G_2 phase	second growth phase of the cell cycle
G-174C	alteration of a guanine to a cytosine at nucleotide position -174 within the promoter region of the IL6 gene
G2161C	alteration of a guanine to a cytosine at nucleotide position 2161 within the Ob-R gene
G-258A	alteration of a guanine to an adenine at nucleotide position -258 within the promoter region of the glucokinase gene
G-30A	alteration of a guanine to an adenine at nucleotide position -30 within the promoter region of the glucokinase gene
G-308A	alteration of a guanine to an adenine at nucleotide position -308 within the promoter region of the TNFα gene
GenBank	GenBank®[1]: United States repository of DNA sequence information
Gln	glutamine
Gln59Leu	alteration of a glutamine to a leucine at amino acid residue 59 within the IPF-1 protein
Glu	glutamic acid
Glu268X	alteration of a glutamic acid to a stop codon at amino acid residue 268 within the HNF4α protein
GLUT	glucose transporter
GLUT 1	glucose transporter 1
GLUT 2	glucose transporter 2
GLUT 4	glucose transporter 4
Gly	glycine
Gly319Ser	alteration of a glycine to a serine at amino acid residue 319 within the HNF1α protein
Gly723del	deletion of a glycine at amino acid residue 723 within the IRS-1 protein
Gly818Arg	alteration of a glycine to an arginine at amino acid residue 818 within the IRS-1 protein
Gly972Arg	alteration of a glycine to an arginine at amino acid residue 972 within the IRS-1 protein
Gly1057Asn	alteration of a glycine to an asparagine at amino acid residue 1057 within the IRS-2 protein
glyc-ox LDL	glycoxidation end product of low density lipoprotein
Grb-2	growth factor receptor binding protein 2
GSK-3	glycogen synthase kinase-3
GSK-3α	glycogen synthase kinase-3 alpha
Gt	genotype
GTP	guanosine triphosphate
GYS1	glycogen synthase
H_A	alternative hypothesis
HDL-C	high density lipoprotein-cholesterol
His	histidine
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HLA-DR4	human leukocyte antigen subtype DR4
HNF1α	hepatocyte nuclear factor 1 alpha
HNF1β	hepatocyte nuclear factor 1 beta
HNF3β	hepatocyte nuclear factor 3 beta
HNF4α	hepatocyte nuclear factor 4 alpha
H_0	null hypothesis

[1] GenBank® is a registered trademark of the National Institutes of Health and Human Services for the Genetic sequence Databank, Bethesda, MD, USA.

List of Abbreviations continued...

HW	Hardy-Weinberg equilibrium
ICA512	islet cell autoantigen 512
ICAM-1	intracellular adhesion molecule-1
i.e.	*id est* meaning "that is to say"
IGT	impaired glucose tolerance
IκB	inhibitor of nuclear factor kappa B
IL6	interleukin 6
Ile	isoleucine
Ile164Thr	alteration of an isoleucine to a threonine at amino acid residue 164 within the APM1 protein
IN	Indiana
insCCG243	insertion of two cytosines and a guanine at nucleotide position 243 within the IPF-1 gene
IPF-1	insulin promoter factor 1
IR	insulin receptor
IRS	insulin receptor substrate
Irs1	murine orthologue of insulin receptor substrate 1
IRS-1 to 4	insulin receptor substrate-1 to 4
ISPK	insulin stimulated protein kinase
JAK	*Janus* kinase
k	number of columns within the contingency table
kb	kilobase pairs
KCl	potassium chloride
$kg.m^{-2}$	kilogram per metre squared: unit of body mass index
KH_2PO_4	potassium phosphate monobasic
LC™	LightCycler™[1]
LC640	LightCycler™ probe which fluoresces at a wavelength of 640 nm
LDL	low-density lipoprotein
LDM	low density microsomes
Leu	leucine
LHON	Leber's hereditary optic neuropathy
LOD_{max}	maximum score of logarithm of odds: measure of linkage
Lys	lysine
m	milli: 10^{-3}
M	molar: moles per litre
µl	microlitre
ml	millilitre
M phase	mitotic phase of the cell cycle
MAPK	mitogen activated protein kinase
MAPK8IP1	mitogen activated protein kinase 8 interacting protein 1
MD	Maryland
MEKK1/2	mitogen activated protein kinase kinase 1/2
MELAS	mitochondrial encephalomyopathy, lactic acidosis and stroke like episodes
Met	methionine
$[MgCl_2]$	concentration of magnesium chloride according to a specific unit
min	minutes
mM	millimolar
mmHg	millimetres of mercury: measure of pressure
$mmol.l^{-1}$	millimole per litre
MODY	maturity onset diabetes of the young
MODY1	locus 1 associated with maturity onset diabetes of the young
MODY2	locus 2 associated with maturity onset diabetes of the young
MODY3	locus 3 associated with maturity onset diabetes of the young
MODY4	locus 4 associated with maturity onset diabetes of the young
MODY5	locus 5 associated with maturity onset diabetes of the young
MODY6	locus 6 associated with maturity onset diabetes of the young
mRNA	messenger ribonucleic acid
mt	mitochondrial
mTOR	mammalian target of rapamycin
n	sample size
n	nano: 10^{-9}

[1] LightCycler ™ is a trademark of Idaho Technology Inc., Salt Lake City, UT, USA.

List of Abbreviations continued...

n_1	affected individuals harbouring the genetic structure
n_2	control individuals harbouring the genetic structure
n_3	affected individuals not harbouring the genetic structure
n_4	control individuals not harbouring the genetic structure
N-terminal	denotes the amino terminus of a polypeptide
NaCl	sodium chloride
Na_2HPO_4	disodium hydrogen phosphate
NCBI	National Centre for Biotechnology Information
NCEP	National Cholesterol Education Program
ND	non differentiable
NEUROD1	neurogenin differentiation factor 1
NFκB	nuclear factor kappa B
ng	nanogram: 10^{-9} gram
n_i	sample size of group i = 1 to 4
NIDDM	non-insulin dependent diabetes mellitus
NIDDM1	non-insulin dependent diabetes mellitus susceptibility locus 1
NIDDM2	non-insulin dependent diabetes mellitus susceptibility locus 2
NIDDM3	non-insulin dependent diabetes mellitus susceptibility locus 3
nm	nanometre
NOSIII	nitric oxide synthase III
NS	non significant
nt	nucleotide
O	observed population distribution
Ob	*obese* gene product or leptin
Ob-R	leptin receptor
OGTT	oral glucose tolerance test
OR	odds ratio
OR_{DL}	global odds ratio calculated via the DerSimonian and Laird method under the random effects model
$OR_{DL\text{-}95\%CI}$	95% confidence interval calculated for OR_{DL}
OR_i	odds ratio calculated to describe the risk within a specific population group
OR_{wolfe}	global odds ratio determined via Wolfe's method under the fixed effects model
$OR_{wolfe\text{-}95\%CI}$	95% confidence interval calculated for OR_{wolfe}
OXPHOS	oxidative phosphorylation
p	short arm of chromosome
p	phosphorylated 3'-end
p	frequency of the 1 allele
p^2	calculated frequency of the 1,1 homozygous genotype
p	pico: 10^{-12}
$p70^{S6K}$	70 kiloDalton S6 kinase
p85	regulatory subunit of phosphoinositide 3'-kinase
p110	catalytic subunit of phosphoinositide 3'-kinase
PBS	phosphate buffered saline [4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4 (pH 7.4), 137 mM NaCl, 2.7 mM KCl]
PCOS	polycystic ovary syndrome
PCR	polymerase chain reaction
PDK-1	3'-phosphoinositide dependent protein kinase-1
PEPCK	phosphoenolpyruvate carboxykinase
pH	indicates acidity: numerically equivalent to the negative logarithm of hydrogen ion concentration expressed in molarity
PH	pleckstrin homology
PI3K	phosphoinositide 3'-kinase
PI-3,4-P_2	phosphatidylinositol-3,4-bisphosphate
PKC	protein kinase C
PLC	phospholipase C
pmol	picomole: 10^{-12} mole
POMC	pro-opiomelanocortin
POWIRS	Profiles of Obese Women with Insulin Resistance Syndrome
PP1	protein phosphatase 1
PPAR α	peroxisome proliferator-activated receptor alpha
PPAR γ	peroxisome proliferator-activated receptor gamma
PPP1R2	protein phosphatase 1 regulatory subunit 2

List of Abbreviations continued...

PPP1R3A	protein phosphatase 1 regulatory subunit 3A
PPRE	peroxisome proliferator-activated receptor response element
P-STAT-3	phosphorylated signal transducers and activators of transcription 3
PTB	phosphotyrosine binding domain
q	long arm of chromosome
q	frequency of the 2 allele
q^2	calculated frequency of the 2,2 homozygous genotype
QTL	quantitative trait loci
Q_{wolfe}	heterogeneity statistic determined via Wolfe's method under the fixed effects model
r	number of rows within the contingency table
R	reverse primer
RACK7	receptor for activated C-kinase type 7
rad	ras associated with diabetes
RAD1	RFMP encompassing the *rad* locus
ras	cellular form of the rat sarcoma proto-oncogene
RFMP	restriction fragment melting polymorphism
RNA	ribonucleic acid
RPNII	ribophorin II
S	Svedberg units
s	second
Ser	serine
Ser465Arg	alteration of a serine to an arginine at amino acid residue 465 within the HNF1β protein
Ser892Gly	alteration of a serine to a glycine at amino acid residue 892 within the IRS-1 protein
Ser/Thr	serine and/or threonine
SH2	*src* homology 2 domain
Shc	SH2 domain-containing oncogenic protein
SH-PTP2	*src* homology 2 domain containing protein tyrosine phosphatase 2
SNPs	single nucleotide polymorphisms
SOS	son of sevenless
src	cellular form of *Rous sarcoma* transforming protein
SSCP	single strand conformation polymorphism
STAT	signal transducers and activators of transcription
T or t	thymine
T-146C	alteration of a thymine to a cytosine at nucleotide position -146 within the upstream region of the HNF4α gene
T1668C	alteration of a thymine to a cytosine at nucleotide position 1668 within the glycogen synthase gene
T_1	maximum melting temperature, at which the fluorescence generated by the probe detecting the 2,2 homozygote peaks
T1D	type 1 diabetes mellitus
T_2	maximum melting temperature, at which the fluorescence generated by the probe detecting the 1,1 homozygote peaks
T2D	type 2 diabetes mellitus
T_a	annealing temperature
Taq	DNA deoxynucleotidyltransferase from *Thermus aquaticus*
TATA	promoter element consisting of the following sequence 5'-TATA-3'
ter	terminal region of a chromosome
TF	transcription factor
Thr	threonine
TLC	Therapeutic Lifestyle Changes diet
TNFα	tumour necrosis factor alpha
tRNA	transfer ribonucleic acid
$tRNA^{Leu(UUR)}$	transfer ribonucleic acid specific for leucine recognising the codon UUR (U = uracil, R = purine)
Trp	tryptophan
Tyr	tyrosine
UCSNP	University of Chicago single nucleotide polymorphism
UCSNP-19	insertion deletion alteration within intron 6 of the calpain 10 gene
UCSNP-43	alteration of a guanine to an adenine within intron 3 of the calpain 10 gene
UCSNP-44	alteration of a thymine to a cytosine within intron 3 of the calpain 10 gene
UCSNP-56	alteration of a guanine to an adenine within intron 6 of the calpain 10 gene that is in linkage disequilibrium with UCSNP-19

List of Abbreviations continued...

UCSNP-63	alteration of a cytosine to a thymine within intron 13 of the calpain 10 gene
USA	United States of America
UT	Utah
Val	valine
Val255Met	alteration of a valine to a methionine at amino acid residue 255 within the HNF4α protein
Val393Ile	alteration of a valine to an isoleucine at amino acid residue 393 within the HNF4α protein
var_i	internal variance within a specific population group
VLDL	very low density lipoprotein
vs	versus
w_{DL}	weight given to a specific population in a meta-analysis under the random effects model which includes the effects of both inter and intra population variability
WHO	World Health Organisation
w_i	weight given to a specific population in a meta-analysis under the fixed effects model
X	fluorescein
X	unknown residue
x g	gravitational acceleration
YAC	yeast artificial chromosome

LIST OF EQUATIONS

LIST OF FIGURES

LIST OF GRAPHS

LIST OF TABLES

List of Tables continued...

List of Tables continued...

List of Tables continued...

ACKNOWLEDGEMENTS

This investigation would not have been possible without the valued contribution and participation of the following people and institutions. I would thus like to thank each of these individuals in turn.

The various patient and control individuals for their willingness to participate in this study without which this investigation would be impossible.

To my supervisor, **Prof. Antonel Olckers** whose undying commitment and invaluable insight has allowed me to experience the true nature of research. She has made such a great impact on my future scientific career that I will never be able to sufficiently thank her. **Prof. Peter Schwarz,** my co-supervisor who has made an amazing opportunity available by inviting me to the Technical University of Dresden in Germany. Furthermore his valuable input into this investigation has meant a great deal to its future success. I thank both of these individuals for allowing me to experience numerous facets of the scientific process.

Dr Annelize van der Merwe whose willingness to share of her expertise as well as her time has truly been an inspiration to me. Her friendship and honesty is a gift for which I will forever be grateful. To my fellow Ph.D. students **Tumi Semete** and **Marco Alessandrini** whose camaraderie and assistance has made my entire postgraduate career a delight. To **Desire Hart,** not only for her administrative support but also for always keeping things in perspective. To **Jake Darby, Michelle Freeman, Dan Isabirye** and **Tharina van Brummelen** for teaching me that in mentoring others, you gain much more than that which you impart. To **Desiré Dalton** whose support showed true commitment. **Martha Sebogoli** and **Maria Maslangu** whose assistance made much more time for me to concentrate on science.

As to the collection of the samples I would like to thank the following people. **Prof Paul Rheeder** and **Dr van Wyk** for access to diabetic clinics at Mamelodi and Kalafong Hospitals, **Dr Loock** and **Dr van Zyl** who allowed access to hypertension clinics as well as **Dr van Deventer** for access to the Madibeng Clinic. I would also especially like to thank **Dr Alta Schutte** for her excellent organisation of the POWIRS1 research program as well

as **Sr Chrissie Lessing** for her superior patient care and troubleshooting skills. **Prof. Lius Perez-Perez** from the National Institute of Endocrinology in Cuba for the collection of both the diabetic and control Cuban cohorts. **Dr Francois van der Westhuizen** for allowing us access to various resources within his laboratory as well as his willingness to assist. I would again like to thank the various postgraduate students of the Centre for Genome Research, mentioned previously, as well as **Christa Mouton, Madeleine Wessels** and **Tharina van Brummelen**. I would also like to convey my great appreciation to **Astrid von Loefellholz** and **Jutta Braun** whose clinical expertise and valuable input into the DIAGEN project have greatly added to its success. Finally I would like to thank **Biochemistry, Nutrition** and **Physiology** for making various resources, equipment as well as laboratory space available.

To **Prof. Schackert** and **Prof. Gräßler** and their laboratories at the Technical University of Dresden, Germany that made valuable contributions in terms of allowing access to equipment. To the various members of the laboratories, namely **Frau Reichmann, Frau Nietsche, Frau von Kannen, Frau Görgens, Babel, Simone** and **Manuela**, for their acceptance of a person who did not speak their language and their willingness to help in spite of this fact. Most importantly I would like to thank **Uta Buro** whose superior technical assistance was implicit in the success of these projects.

The Centre for Genome Research for making the infrastructure available wherein a high quality of science could be achieved as well as for its financial assistance in the form of a post-graduate bursary. **DNAbiotec (Pty) Ltd,** which made various resources available without which this study would not have been completed. **Fifth Dimension Technologies (5DT)** for various resources and more specifically **Paul Olckers** who made his vast experience freely available. The **National Research Foundation** for financial assistance in the form of an NRF Prestigious Scholarship for Doctoral Study.

The **North-West University (Potchefstroom Campus)** for making this unique environment possible wherein academic and commercial entities can work hand in hand in the more holistic education of students. Specifically I would like to thank the Dean of the Faculty of Natural Sciences namely **Prof Daan van Wyk** as well as the Head of the Research Focus Area of Preventative and Therapeutic intervention, **Prof Estie Vorster** for their support which has added a great deal of value, not only to this investigation but the greater research programs within the **Centre for Genome Research**. The **Research**

Focus Area 9.1 furthermore made strategic funding available to the CGR. I would also greatly like to thank the staff at the **Ferdinand Postma library**

My extended family and friends for their love, understanding and support. All my **Grandparents** who gave me a place to stay, someone to depend on, and more love than any one person deserves. To **my parents** and **my brother** whose love, support, encouragement and empathy have helped me to achieve my goals and whom I will never be able to thank sufficiently for the wonderful opportunities they have made available to me. Finally to the **Lord** who has always been my strength and by whose grace I have been truly blessed.

CHAPTER ONE

INTRODUCTION

Type 2 diabetes mellitus (T2D) is currently the fastest growing epidemic of non-communicable disorders worldwide (King *et al.*, 1998). The associated epidemic of obesity is similarly experiencing a concomitant increase (Formiguera and Cantón, 2004). The major reason given for these increases, is the exposure of non-Western countries to the so-called Westernised diet of high fat and simple carbohydrates, when historically the populations of the so-called developing countries maintained a diet of complex carbohydrates and protein and are thus not capable of efficient metabolism of the aforementioned molecules (Nakanishi *et al.*, 2004). Collectively these disorders comprise the metabolic syndrome, which originates from resistance to insulin (Reavan, 1988). The strain of these preventable disorders weighs heavily on medical systems globally especially when it is taken into consideration that health care institutions are currently affected by the increase in infection rates of the human immunodeficiency virus (HIV) and acquired immune deficiency syndrome or AIDS (Benatar, 2004).

The group of disorders which constitute T2D are defined by high levels of glucose within the circulatory system (World Health Organisation (WHO) Consortium, 1999). This is caused by defects within the cellular metabolism, generally brought about by defects in cellular signalling. The major signalling pathways involved are discussed in greater detail in Chapter Two however it is the intracellular effects of insulin (Rossetti and Giaccari, 1990) and such adipocytokines as leptin and adiponectin that are generally affected (Cohen *et al.*, 1996; Weyer *et al.*, 2001).

Diabetes mellitus (DM) is an encompassing term and includes numerous subdivisions according to the origins of the major symptom of hyperglycaemia which are discussed in Chapter Two. The origins of T2D are complex in nature and generally caused by an increased resistance to insulin (Polonsky *et al.*, 1996) resulting in eventual beta (β) cell failure. Factors affecting the expression of T2D have been determined at the genetic (Barnett *et al.*, 1981) and environmental (Bergman *et al.*, 1981) levels. These factors work in parallel to induce disease susceptibility in an individual. These are however not the only modifiers of disease risk. Obesity is a major cause of the metabolic syndrome and by

association T2D due to the effect it has on insulin sensitivity (Kolterman *et al.*, 1980). The increased risk imbued via this metabolic state is described in Chapter Two.

Within Chapter Two the various biochemical and clinical risk factors are broadly discussed, however the delineation of the genetic risk factors involved in T2D is required, which is presented in Chapter Three. This chapter includes an overview of the candidate genes for T2D susceptibility followed by the role that murine models have fulfilled in the elucidation of the genetic aetiology of this disorder. Finally the various genes that have been genetically linked to T2D especially calpain 10 (CAPN10) and adiponectin (APM1), is reviewed.

In the era of genomics, the importance of single nucleotide polymorphisms (SNPs) within the non-coding regions of genes is apparent (Gray *et al.*, 2000). It is therefore not surprising that the various susceptibility loci discussed have been determined to harbour SNPs which are associated with the T2D phenotype (Horikawa *et al.*, 2000; Vasseur *et al.*, 2002).

Non-insulin dependent diabetes mellitus susceptibility locus 1 (NIDDM1) was localised to the long arm (q) of chromosome 2 within the Mexican American population. Positional cloning of this locus resulted in the elucidation of the CAPN10 gene, which harboured various SNPs associated with disease risk. This gene and its association to T2D susceptibility are reviewed within Section 3.3.1.

A locus on chromosome 3q27 was linked to various symptoms of the metabolic syndrome (Kissebah *et al.*, 2000). The gene encoding an adipocytokine termed adiponectin was localised to the same region (Saito *et al.*, 1999). Low levels of this gene product were associated with an increased risk towards T2D. Furthermore, hypoadiponectinaemia was also present in obese individuals indicating a possible mechanism of action (Spranger *et al.*, 2003). Alterations associated with T2D were determined however the risk loci were different within the European (Vasseur *et al.*, 2002) and Japanese (Hara *et al.*, 2002) populations.

A pilot study investigating the presence of the at risk loci within CAPN10 and APM1 undertaken within the greater research program resulted in the elucidation that the black South African cohort investigated had different allele and haplotype frequencies upon comparison to the non-African populations (Towers, 2002). This was however expected as

this population is part of the larger African macrohaplogroup L lineage, which harbours the greatest level of genetic variation, due to it being the ancestral population from which all others arose (Chen *et al.*, 1995). This finding also led to the hypothesis that any treatment strategy developed according to the at-risk alleles within the developed countries would not be applicable to the individuals within the developing countries. Determination of whether the associations present within the developed countries are also responsible for disease risk within developing countries is therefore required.

The largest increase in T2D susceptibility over the next 20 years will be within the developing countries (King *et al.*, 1998) and thus such an investigation is essential. If it is taken into consideration that most exploration of T2D susceptibility concentrates on so-called 'First World' countries it is required of scientists within developing countries to elucidate the genetic aetiology of the disease within their own populations. This investigation was undertaken to determine if the reported risk factors are similarly associated with disease risk within the black South African and Cuban populations via comparison to various non-African populations utilising a case-control study design.

The criteria utilised for collection of these individuals as well as the methods of obtaining DNA samples are described in Chapter Four. Screening of four SNPs within CAPN10 and three SNPs within the APM1 gene was achieved via real time polymerase chain reaction and melting curve based analyses utilising the LightCycler™[1] technologies. Haplotype assignment as well as the various statistical analyses utilised to determine significance are also discussed within Chapter Four.

The findings of this investigation are presented in Chapter Five and the conclusions determined are outlined in Chapter Six. Within these chapters the significance of ancestral lineages to the investigation of T2D susceptibility is highlighted as well as the necessity for the development of new molecular tools for the elucidation of mechanisms of disease action. Ultimately, the development of affordable yet efficient system biology strategies is required to elucidate and eventually eradicate T2D.

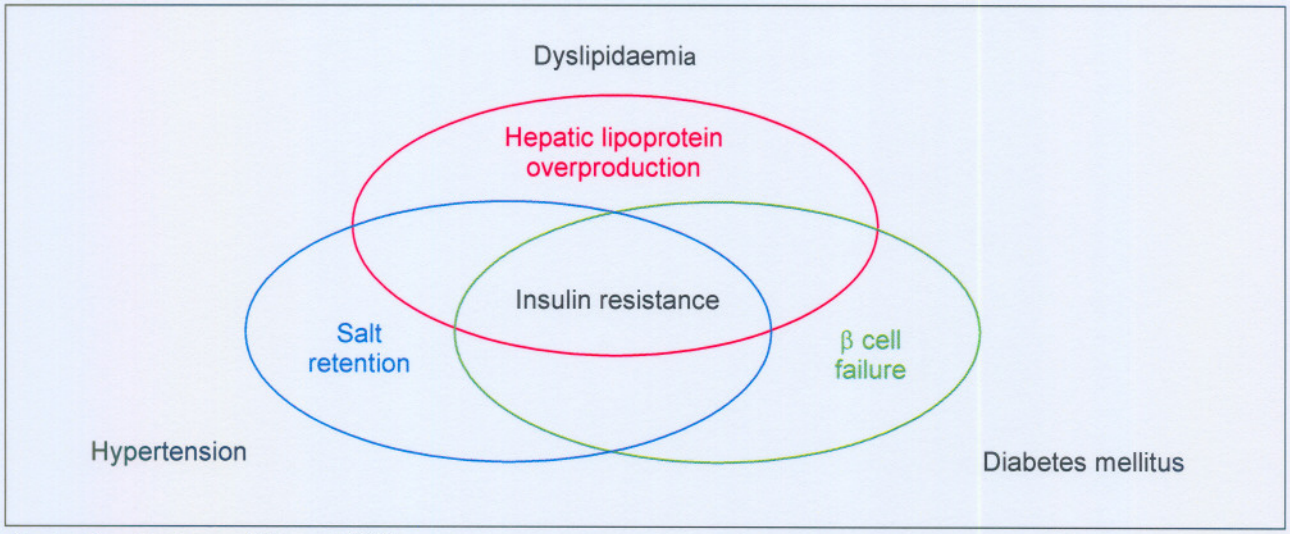
[1] LightCycler™ is a trademark of Idaho Technology Inc., Salt Lake City, UT, USA.

CHAPTER TWO

THE CLINICAL AND BIOCHEMICAL ASPECTS OF TYPE TWO DIABETES MELLITUS

The metabolic syndrome is a group of abnormalities of the metabolism with the primary fault occurring in an individual's sensitivity towards insulin, resulting in the major symptoms of hyperinsulinaemia and insulin resistance (Reavan, 1988). Various derived complications such as hypertension, coronary artery disease (CAD) and T2D generally follow this initial defect as depicted in Figure 2.1.

Figure 2.1: Diagrammatic representation of the causes of the various symptoms originating from the metabolic syndrome



Adapted from Hughes and Aitman (2004).

The following diagnostic criteria for this cluster of disorders were initially defined with regard to the treatment of patients suffering from high levels of low-density lipoprotein (LDL) cholesterol. The metabolic syndrome is, however, defined in accordance with the criteria specified by the Adult Treatment Panel III (ATPIII) of the National Cholesterol Education Program (NCEP) of the United States of America (USA). The threshold values of these criteria are depicted in Table 2.1.

Table 2.1: Diagnostic criteria of the metabolic syndrome

Diagnostic criterion	Men	Women
waist circumference	> 102 cm	> 88 cm
HDL-C levels	> 1.0 mmol.l^{-1}	> 1.3 mmol.l^{-1}
triglyceride levels	> 1.7 mmol.l^{-1}	
FPG	> 6.1 mmol.l^{-1}	
blood pressure	> 130/85 mmHg	

HDL-C = high density lipoprotein-cholesterol; FPG = fasting plasma glucose; mmHg = millimetres of mercury; mmol.l^{-1} = millimole per litre; cm = centimetre; Adapted from the Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (2001).

T2D is therefore only one of the numerous outcomes of insulin resistance. However, because this specific group of disorders are rapidly reaching epidemic proportions (King *et al.*, 1998) it is important to determine the various genetic, environmental and biochemical factors, which result in the specific manifestation of T2D prior to the development of effective treatment strategies.

2.1 DIABETES MELLITUS

DM is a term utilised to describe a phenotypically heterogeneous group of metabolic disorders having numerous aetiologies characterised by chronic hyperglycaemia and disturbances in the metabolism of lipids, protein and specifically carbohydrates due to insulin loss, insulin insensitivity or both. This disruption may lead to the dysfunction and failure of the different organ systems (WHO Consortium, 1999).

Approximately 4 percent (%) of the worldwide population is affected by this group of disorders, as determined in 1995, and it has been predicted that the prevalence will increase to 5.4% by the year 2025 (King *et al.*, 1998). Although more prevalent in developed countries, currently the main cause of this increase will be due to the increased prevalence of diabetes in developing countries which will be an increase of 3.3% to 4.9% (King *et al.*, 1998). Numerically this translates to an increase from 84 million to 228 million people affected by the year 2025.

Of the various developing countries, Sub Saharan Africa will account for the third largest increase of people affected by this group of disorders i.e. an increase of 185% which translates to 5 million newly affected individuals (King *et al.*, 1998). This information was, however, only based on a few reports and therefore may represent an underestimation. The largest increase within this group, it is hypothesised, will occur in individuals between

the ages of 20 and 64 (King *et al.*, 1998). Individuals will be affected for longer periods of time, placing greater strain on the limited medical resources. Given that the various health care systems operating within this region are already strained by the growing burden of HIV and AIDS (Benatar, 2004), the additional burden of T2D can have devastating effects. It is thus imperative to undertake the elucidation of the predisposing factors towards this group of disorders in the African population in order to implement effective prevention strategies.

The classification of this group of disorders is complex and has undergone numerous revisions. The initial attempt was undertaken by Himsworth in 1936 where the original causation of the disorders was utilised as the defining characteristic. In this nomenclature the group was divided into insulin sensitive and insulin insensitive diabetes mellitus. Review of this rather simplistic classification has lead to the classification of these disorders based on the aetiology of the hyperglycaemia, as well as the clinical staging of the individual (WHO Consortium, 1999).

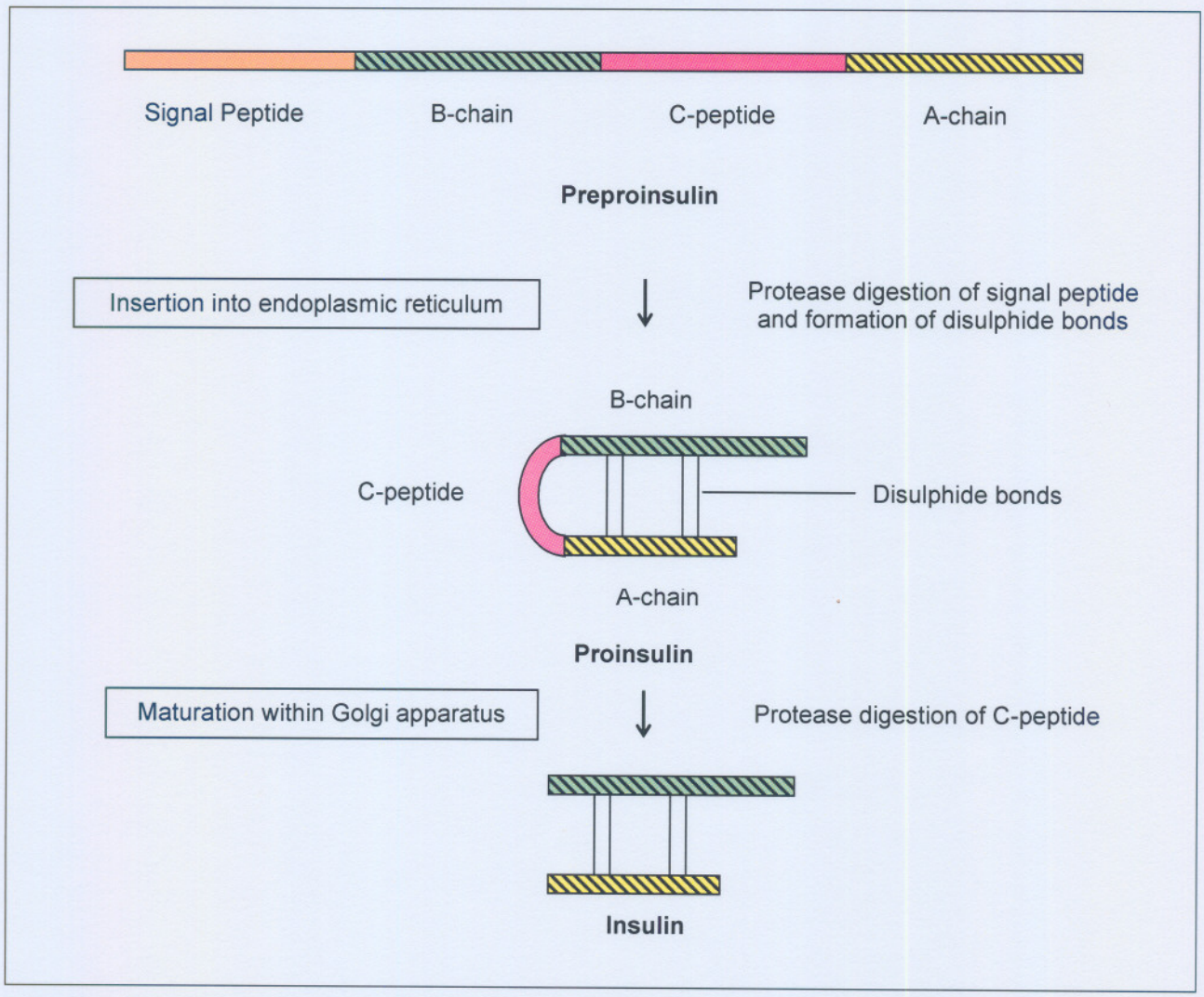
2.1.1 Insulin

Any discussion of the aetiology of this diverse group of disorders initially requires an overview of the hormone insulin and its role in cellular metabolism. Insulin has both excitatory and inhibitory activities and acts as a homeostatic control between gluconeogenesis and cellular oxidation as well as other metabolic functions (Rossetti and Giaccari, 1990). Insulin first causes a decrease in hepatic glucose production. Secondly, it increases the uptake of glucose by the various tissues, mainly skeletal muscle, therefore resulting in the decreased plasma levels of this saccharide (Brown *et al.*, 1978; DeFronzo *et al.*, 1981). Its major function, however, is prevention of the utilisation of alternative sources of energy e.g. gluconeogenesis, glycogenolysis (Brown *et al.*, 1978), lipolysis (Thomas *et al.*, 1979), ketogenesis (Randle, 1985) and proteolysis when glucose is present.

The β cells of the pancreas produce insulin in the form of a preprohormone. Preproinsulin consists of three major subunits namely the A, B and C-peptides, as well as a signal peptide directing it towards the endoplasmic reticulum as presented in Figure 2.2. Following insertion into the membrane the signal peptide is proteolytically cleaved to produce proinsulin. Upon activation, the C-peptide is cleaved to produce active insulin

within the cell, which is subsequently released into the peripheral circulatory system. The A and B peptide chains remain attached via disulphide bonds (Narang *et al.*, 1984).

Figure 2.2: Diagrammatic representation of the synthesis of the insulin hormone from preproinsulin



Insulin is released in a pulsatile manner i.e. it is released periodically instead of continuously (Hansen *et al.*, 1982). It has been determined that after the ingestion of nutrients (especially carbohydrates) following a period of fasting, insulin levels fluctuate with a periodicity of a mean value of twelve minutes. Hansen *et al.* also presented data indicating that the ensuing oscillation of the plasma glucose level was synchronous with the oscillation of the plasma insulin levels with a period of two minutes between each. The insulin therefore directly induces a decrease in the plasma glucose level. Further investigation has determined that it is the relative level of hyperglycaemia following a meal that induces the release of insulin in such a manner (Kingston and Skoog, 1986).

2.1.1.1 Signalling pathways of insulin

The insulin receptor (IR) is a heterotetrameric glycoprotein consisting of two alpha (α) and two β subunits. A single polypeptide of 1,355 amino acid residues is translated from the messenger ribonucleic acid (mRNA) derived from the IR gene. This polypeptide includes both the α (735 amino acids) and β (620 amino acids) subunits, which are only proteolytically cleaved following translation. The α subunit harbours a cysteine rich domain which is involved in the cross linking of the various receptor molecules. A function for the amino (N)-terminal domain has not yet been assigned, whilst it has been hypothesised that the carboxy (C)-terminal region is involved in hormone binding. The β subunit, however, consists of a transmembrane domain and a cytoplasmic domain responsible for its tyrosine kinase ability. It has also been determined that the extracellular region of the β subunit harbours four glycosylation sites while the α subunit contains 18 residues that can be glycosylated (Ebina *et al.*, 1985).

Phosphorylation of the tyrosine residues at positions 1,158, 1,162 and 1,163 within the receptor, induces a remarkable alteration in the conformation of the activation loop, thus allowing for access of the binding domain by adenosine triphosphate (ATP) and the various protein substrates. The phosphotyrosine at position 1,162 is present within the active site and binds to the tyrosine to be phosphorylated in the substrate protein. Phosphorylation is unlikely to occur in an associative manner as the gamma (γ) phosphate of the nucleotide triphosphate is incorrectly positioned with regard to the substrate (Hubbard, 1997). An important aspect to consider is that the receptor retains its activity even after the dissociation of insulin. The receptor is only repressed by its dephosphorylation (Rosen *et al.,* 1983) thus allowing for various feedback mechanisms to control its activity.

2.1.1.1.1 The role of the insulin receptor substrate (IRS) proteins

Upon activation of the receptor via tyrosine phosphorylation the insulin-receptor complex sequesters various signal transduction proteins. The insulin receptor substrate (IRS) isoforms are a major group of proteins induced by the activation of the insulin receptor (Sun *et al.*, 1991; Withers *et al.*, 1998; Yu *et al.*, 1998). The various IRS isoforms in turn harbour a pleckstrin homology (PH) domain, which localises the specific protein towards the plasma membrane. PH domains are integral to the activation of insulin receptor substrate 1 (IRS-1) and insulin receptor substrate 2 (IRS-2) as these domains are

responsible for membrane localisation of the IRS proteins via phospholipid interaction. This in turn allows for activation of the aforementioned proteins due to the close proximity to the insulin receptor (Voliovitch *et al.*, 1995).

Attachment to the receptor is achieved via the phosphotyrosine binding (PTB) domain of the IRS isoforms. The PTB domains of the IRS proteins form a complex with the juxtamembrane region of the insulin receptor (Wolf *et al.*, 1995). PTB is a seven-stranded β sandwich capped by a long α helix. The phosphopeptide sequence within the insulin receptor fills an L shaped cleft in the surface formed by the β5 sheet, the C-terminal helix and the B10 turn connecting β4 and β5. This sequence forms a β sheet that undergoes hydrogen bonding with the β5 sheet in an anti-parallel fashion. The phosphotyrosine at position 1,162 within the insulin receptor, co-ordinates with two arginines (Arg), one of which is in close contact with the aromatic ring of the tyrosine. Various hydrogen and ionic bonds ensure co-ordination of the oxygen present in the phosphate thus ensuring that only phosphorylated tyrosine is bound. Various other interactions ensure specificity (Eck *et al.*, 1996). This results in the phosphorylation of important tyrosine residues within the IRS isoforms, which in turn allows for the docking of various downstream effector proteins.

Upon analysis of double and triple heterozygous knockout mice for the insulin receptor, IRS-1 as well as IRS-2 it was discovered that those lacking in the IR/IRS-1 pathway presented with severe insulin resistance in skeletal muscle and liver. IR/IRS-2 knockout mice presented with severe insulin resistance in the liver and mild resistance in skeletal muscle. This indicates that these substrates may have different effects in different tissues (Kido *et al.*, 2000). Via analysis of certain knockout murine models it has been determined that the IRS isoforms are not only responsible for insulin sensitivity but are also necessary for the proper production and secretion of β cells. By knocking out the IRS-1 gene it was determined that the insulin content of the β cells was decreased and the glucose stimulated insulin secretion was severely impaired (Kulkarni *et al.*, 1999). This fact provides further evidence for the possibility that the IRS proteins may have tissue specific effects.

One of the major effector pathways activated via IRS-1 and -2 is that of phosphoinositide 3'-kinase or PI3K (Sun *et al.*, 1991; Withers *et al.*, 1998; Yu *et al.*, 1998). This is achieved via the formation of complexes between PI3K, IRS and the insulin receptor upon stimulation by insulin (Backer *et al.*, 1993). Signalling complexes may consist of the

tyrosyl-phosphorylated IRS bound to the insulin receptor, which in turn induces relocalisation of PI3K from the cytosol to the plasma membrane. Alternatively the activated IRS may be released into the cytosol whereupon it binds to the PI3K and induces its activity. The two possible mechanisms of signal transduction may ultimately have differential effects depending on the cytosolic localisation of PI3K upon activation (Backer *et al.*, 1993).

The regulatory p85 subunit of the PI3K enzyme complex is bound to the catalytic p110 subunit via an interim domain between two transforming protein of *Rous sarcoma* (*src*) homology 2 (SH2) domains. This association has been determined to increase the stability of the catalytic (p110) subunit, however, in its unbound form the p85 subunit inhibits the activity of the p110 subunit (Yu *et al.*, 1998). The tyrosine activated IRS-1 and -2 in turn interact with the p85 subunit through the recognition of the phosphorylated tyrosine-methionine-unknown residue-methionine (Tyr-Met-X-Met) motifs via the aforementioned SH2 domains (Myers *et al.*, 1992). Binding results in the activation of the catalytic capabilities of the p110 subunit (Hu *et al.*, 1993). This has been hypothesised to be due to a loss of the inhibitory function of the p85 subunit induced via the binding of the phosphorylated IRS proteins (Yu *et al.*, 1998). Further evidence for this hypothesis has been determined in the investigation of diabetic murine models whereby reduced levels of p85 improved insulin signalling as well as the overall diabetic phenotype (Mauvais-Jarvis *et al.*, 2002).

Analysis of *Xenopus* oocytes resulted in the determination that the PI3K pathway is implicit to glucose transport (Gould *et al.*, 1994). Protein kinase B (Akt) may be a possible effector molecule by which this is achieved. Activation of PI3K, through insulin stimulation, induces the activation of the serine and/or threonine (Ser/Thr) kinase Akt (Brozinick and Birnbaum, 1998) via binding of phosphatidylinositol-3, 4-bisphosphate (PI-3,4-P_2) to the PH domain, thus directing the kinase to the plasma membrane. This in turn increases dimerisation ability (Franke *et al.*, 1997). The protein can also be activated via direct phosphorylation by PI3K, which allows for the downstream activation of 70 kiloDalton S6 kinase or $p70^{S6K}$ (Kohn *et al.*, 1996a). Activated Akt increases glucose uptake via two pathways. Firstly, it induces the translocation of glucose transporter (GLUT) 4 to the plasma membrane and secondly, it increases the expression of GLUT 1 (Kohn *et al.*, 1996b). The activated kinase has a further function in that it increases lipidogenesis and glycogen synthesis. Kohn *et al.* in 1996a produced an Akt construct, which was directed towards the plasma membrane. In the case of cells expressing this form of the kinase there was an increase in lipidogenesis

but not in glycogen synthesis upon insulin treatment, however, both pathways were upregulated in control cell lines. Taking into account the fact that there are two pathways by which Akt is activated it is thus plausible that there are separate functions for the plasma membrane associated and non-associated forms of Akt. For example the plasma membrane associated form may cause an increase in lipidogenesis, whereas the non-associated form induces glycogen synthesis.

IRS-1 phosphorylation is also integral to the initiation of 2'-deoxyribonucleic acid (DNA) synthesis as well as the progression into the second growth (G_2) and mitotic (M) phases of the cell cycle (Valverde *et al.*, 2001). Tyrosine phosphorylated IRS-1 achieves this via its interaction with growth factor receptor binding protein 2 (Grb-2). A further effect of the activated IRS-1 is to bind and activate the SH2 domain-containing oncogenic protein also termed the Shc protein (Valverde *et al.*, 2001). Both IRS-1 and Shc, in turn associate with Grb-2 via its SH2 domain (Skolnik *et al.*, 1993). Grb-2 does not undergo a phosphorylation event and must therefore be activated via an induced conformational change. Grb-2 activates the mitogenic pathway via induction of the cellular form of the rat sarcoma proto-oncogene (ras). This is achieved by the association of Grb-2 and a guanine nucleotide releasing factor known as son of sevenless (SOS) in response to insulin treatment (Skolnik *et al.*, 1993). It is hypothesised that this interaction in association with the uptake of Grb-2/SOS complex by IRS-1 and Shc results in the activation of ras. This is achieved by the translocation of the cytoplasmic Grb-2/SOS complex to the plasma membrane thus facilitating interaction with the ras protein. SOS has been determined to increase the guanosine triphosphate (GTP) binding ability of ras threefold (Baltensperger *et al.*, 1993). Ras in turn further mediates the effect by the activation of mitogen activated protein kinase kinase (MEKK1/2) and p42-p44 mitogen activated protein kinase or MAPK (Valverde *et al.*, 2001). This process is presented in a model of insulin signal transduction in Figure 6.1 and 6.2.

The *src* homology 2 domain containing protein tyrosine phosphatase 2 (SH-PTP2) has also been determined to be involved in this pathway, however, the mechanism is as yet unclear (Noguchi *et al.*, 1994). The molecule binds IRS-1 upon insulin treatment, causing dephosphorylation of IRS-1 at a specific tyrosine residue. This in turn causes activation of ras via an undefined pathway. It was hypothesised by Naguchi *et al.* that the activated SH-PTP2 may have an upstream effect which mediates the Grb-2/SOS pathway as these proteins themselves are not affected by the overexpression of a catalytically inactive SH-PTP2 (Noguchi *et al.*, 1994).

Disruption of IRS-1 in mice, results in a 40-60% decrease in birth weight as compared to wild type individuals (Araki *et al.*, 1994). This effect is maintained throughout the lifetime of the IRS-1 deficient mouse. The IRS-1 knockout mice furthermore presented with impaired glucose tolerance and the response to both insulin and insulin growth factor-1 was severely reduced. Upon analysis of PI3K activity it was determined that, although IRS-1 associated activation of this protein was impaired, the overall activation of PI3K was not greatly modified (Araki *et al.*, 1994). It was subsequently determined that IRS-2 fulfilled this role (Patti *et al.*, 1995).

Although IRS-2 activates PI3K in a similar fashion to IRS-1 it has been determined that the further downstream signalling is different with regard to this protein (Arribas *et al.*, 2003). Induction of PI3K activity results in the activation of a protein kinase named 3'-phosphoinositide dependent protein kinase-1 (PDK-1). It acts via the phosphorylation of the protein kinase C (PKC) zeta (ζ) isoform (Le Good *et al.*, 1998). Insulin thus induces rapid phosphorylation of PKC ζ through the action of PDK-1 via PI3K activation. PDK-1 associates with most PKC isoforms and is directly responsible for activation of the delta (δ) and ζ forms (Le Good *et al.*, 1998). Over-expression of PKC ζ increases basal and insulin stimulated glucose transport (Bandyopadhyay *et al.*, 1997). This is achieved by inducing the movement of GLUT 1 and 4 from the microsomes to the plasma membrane (Etgen *et al.*, 1999). Glucose sensitisation of the cell is thus achieved via the IRS-2/ PI3K/PKC delta pathway.

As stated previously, defects in insulin secretion are an aggravating factor in susceptibility towards T2D upon insulin resistance. It is therefore not surprising that alterations within the IRS-2 protein have effects at both of these levels. Its role in insulin sensitivity was discussed in the previous paragraph. Mice deficient in IRS-2 have decreased β cell mass as compared to mice harbouring functional IRS-2 (Withers *et al.*, 1998). Upon expression analysis of this protein in wild type mice it was determined that the protein was associated with insulin in β cells as well as being present in the β cell progenitor tissue known as ductal epithelium. In the IRS-2 deficient mice no such expression was determined. This therefore led to the hypothesis that IRS-2 may be integral in the differentiation and maintenance of β cells, however, further analysis is required in order to validate this postulate (Withers *et al.*, 1998).

If the fact that IRS-2 may be important for β cell differentiation is taken into consideration it becomes evident that these two IRS isoforms may have different roles in signalling. IRS-1 may be preferentially responsible for the transfer of the growth induction signal as indicated by its necessity in cell cycle progression (Valverde *et al.*, 2001), and somatic growth (Araki *et al.*, 1994). Alternatively IRS-2 may be more specifically involved in the classic transduction of the insulin signal to ensure correct glucose transport (Etgen *et al.*, 1999) as well as proliferation of β cells (Withers *et al.*, 1998). It therefore seems more plausible that defects in IRS-2 are responsible for the diabetic phenotype. This hypothesis is discussed in greater detail in Chapter Three.

2.1.1.1.2 Stimulation of glycogen synthesis

There are two major processes by which insulin increases the activity of glycogen synthase. The first is mediated via the direct dephosphorylation of glycogen synthase (Dent *et al.*, 1990) whereas the second acts through the inactivation of glycogen synthase kinase-3 or GSK-3 (Embi *et al.*, 1980). Upon activation of the signal transduction network in response to insulin as previously discussed in Section 2.1.1.1, the insulin stimulated protein kinase (ISPK) is phosphorylated and thus activated. In turn this molecule is able to phosphorylate a specific Ser residue within the glycogen bound subunit of protein phosphatase 1 (PP1) which increases its phosphatase ability. Increased PP1 activity results in decreased phosphorylation of glycogen synthase, which allows for increased production of glycogen (Dent *et al.*, 1990).

As stated in the previous paragraph the second method of glycogen synthase activity is via repression of GSK-3. GSK-3 is activated via tyrosine phosphorylation at amino acid residue 216. Unlike the various other pathway effector molecules that have been discussed GSK-3 is active within resting cells (Hughes *et al.*, 1993). The β isoform of GSK-3 phosphorylates eukaryotic initiation factor 2B (eIF2B) at specific Ser and Thr residues thus inactivating it. eIF2B is phosphorylated specifically at the p84 or (epsilon) ε subunit. It is involved in peptide chain initiation via ensuring the uptake of the initiator methionine transfer ribonucleic acid (tRNA) by the ribosome (Rhoads *et al.*, 1993). The ability of this specific kinase to control the function of this protein is lost upon treatment with insulin. This inactivation is rapid, i.e. <10 minutes. GSK-3 also phosphorylates glycogen synthase (Embi *et al.*, 1980) which also presents with decreased phosphorylation upon insulin treatment. This reduced activity is brought about by phosphorylation of certain Ser and Thr residues within GSK-3 via an insulin activated protein kinase. Inactivation of

GSK-3 resulted in decreased phosphorylation of glycogen synthase and eIF2B, thus increasing the function of these proteins (Welsh and Proud, 1993). Over-expression of PKC α, β 1 and 2 isoforms results in the inhibition of insulin effects on glycogen synthesis. Thus insulin-stimulated recruitment of these PKC isoforms may inhibit glycogen synthesis via phosphorylation of GSK-3. GSK-3 is a central effector molecule for various pathways and can be activated via the action of the endothelium-cell differentiation (Edg) family through a PKC-dependent pathway (Fang *et al.*, 2002). It is also responsive towards certain G-protein coupled receptors via protein kinase A dependent pathway through alterations in the cellular cyclic adenosine monophosphate (cAMP) levels (Fang *et al.*, 2000).

2.1.1.2 Negative feedback mechanisms of insulin

The strict regulation of the insulin signalling network requires that certain feedback mechanisms are present which negatively control the cascade. A major example of this is the role of the ε and ζ isoforms of the 14-3-3 proteins (Ogihara *et al.*, 1997). It has been determined that these proteins are able to bind IRS-1 at Ser phosphorylated residues at amino acid positions 270, 374 and 641. Position 270 is within the PTB domain of IRS-1 and when phosphorylated it causes decreased interaction with the insulin receptor, thus impairing Tyr phosphorylation. The interaction between IRS-1 and the 14-3-3 protein does not increase proportionally with insulin stimulation, however, it was determined that treatment with a potent Ser/Thr phosphatase inhibitor known as okadaic acid resulted in increased affinity (Ogihara *et al.*, 1997). It was thus hypothesised that increased Ser/Thr phosphorylation of IRS-1, as brought about by the activation of various terminal members of the insulin signalling pathway, is responsible for increased binding of 14-3-3 protein which results in decreased levels of Tyr phosphorylation thus inhibiting the signal cascade (Ogihara *et al.*, 1997).

Another pathway controlled via insulin stimulation is known as the mammalian target of rapamycin (mTOR) pathway. This pathway induces the movement of IRS-1 from the low density microsomes (LDM) to the cytosol, which ultimately results in the proteasomal degradation of this protein. The mTOR pathway is activated via PI3K activity. It is important in the insulin-induced decrease in the levels of IRS-1 over time via increasing proteasomal degradation of the aforementioned protein. It also regulates the Ser/Thr phosphorylation of IRS-2 as well as the redistribution of this protein from LDM to the

cytosol. However, IRS-2 is not degraded over time via this pathway (Haruta *et al.*, 2000; Takano *et al.*, 2001).

Ultimately the insulin-insulin receptor complex is internalised, thereby inducing the movement of the glucose transport activity from the Golgi apparatus or a similar micro vesicle to the plasma membrane (Suzuki and Kono, 1980). This in turn results in an increase in the membrane permeability towards glucose.

2.1.2 Obesity and its role in disease pathogenesis

Obesity is one of the major risk factors for T2D but its involvement in disease pathogenesis is as yet not fully understood (Kolterman *et al.*, 1980). This physiological state is due to caloric intake being greater than energy expenditure. Although belief to the contrary is common it has been ascertained that obesity is genetically controlled (Friedman, 2000). Various cytokines have been determined to be associated with obesity and T2D, and will be discussed in greater detail (Couce *et al.*, 1997; Spranger *et al.*, 2003).

2.1.2.1 Leptin

The leptin protein was determined to be involved in the regulation of body weight by inducing hypophagia, increased activity and preferential utilisation of lipids as energy sources (Pelleymounter *et al.*, 1995). Furthermore leptin is secreted as a soluble protein by adipocytes to ensure that lipid accumulation only occurs in these specific cells (Shimabukuro *et al.*, 1998).

The protein itself is globular in nature, consisting mainly of polar hydrophillic amino acids. Increased adiposity results in the increased expression of this protein (Zhang *et al.*, 1994). This in turn causes reduced feeding and increased catabolism of free fatty acids (FFAs). In certain obese patients, however, it has been determined that, either due to leptin or its receptor being inactivated, the signal is not transduced to the cell, therefore resulting in hyperphagia and obesity (Surwit *et al.*, 1997).

The receptor contains a signal sequence for secretion of the protein, an extracellular domain similar to the glycoprotein 130 domain of interleukin-6, a transmembrane domain and a cytoplasmic region containing a box 1 *Janus* kinase (JAK) interaction sequence hence this proteins' signal transduction ability (Tartaglia *et al.*, 1995). The major sites of

expression are mainly in the lung and kidney, and various other regions at lower levels such as the brain i.e. specifically in the choroid plexus and hypothalamus. As the hypothalamus is involved in energy homeostasis, this is a possible example of direct signal transduction controlling body weight (Couce *et al.*, 1997).

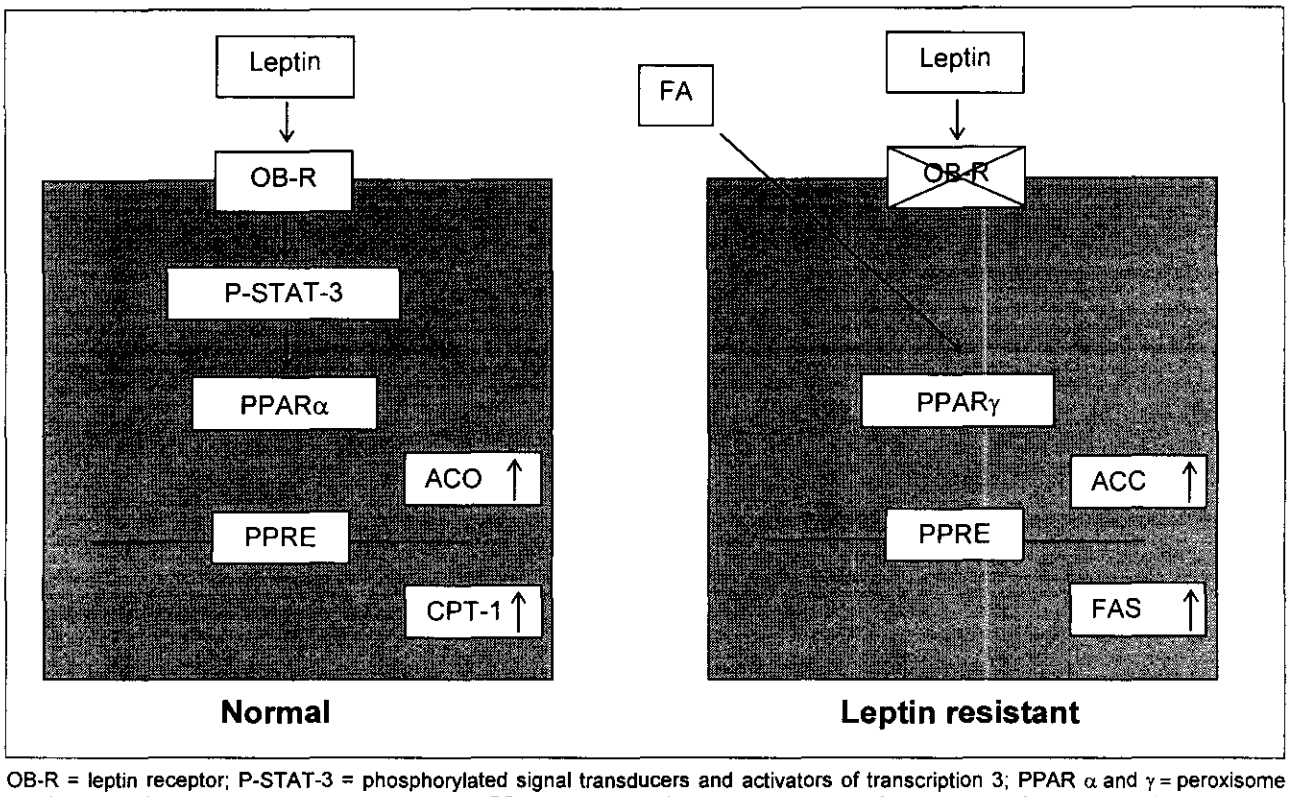
Signal transducers and activators of transcription (STAT) 3 activation is fundamental to the effects that leptin has on the regulation of food intake as well as that of energy utilisation (Bates *et al.*, 2003). Binding of leptin to its receptor in cells of the hypothalamus as well as various types of non-adipose tissues, results in the activation of the JAK family of signal transducers. JAK transducers in turn specifically phosphorylate STAT-3, now abbreviated as P-STAT-3 (Ghilardi *et al.*, 1996). In the hypothalamic cells activation of this pathway results in the increased expression of pro-opiomelanocortin (POMC), which is hypothesised to induce hypophagic effects (Bates *et al.*, 2003). However, as depicted in Figure 2.3, within the non-adipocytes the phosphorylated STAT proteins stimulate the production of the transcription factor (TF) peroxisome proliferator activated receptor α (PPARα). This TF induces expression of proteins such as acyl-coenzyme A (CoA) oxidase (ACO) and carnitine palmitoyl transferase-1 (CPT-1). These proteins increase oxidation of triglycerides thus preventing the accumulation of these lipids (Unger *et al.*, 1999).

The leptin protein also has an effect on insulin and its activity in the cell (Segal *et al.*, 1996). This is achieved by the dephosphorylation of the major substrate of the insulin receptor i.e. IRS-1. Phosphorylation of IRS-1 is necessary for the activation of various signal transduction pathways involved in the induction of insulin-mediated activities, whereas its dephosphorylation results in loss of signal and hence the resistance toward insulin. This in turn may result in a T2D phenotype (Cohen *et al.*, 1996). Furthermore it has been determined that a lack of IRS-2 activity in the hypothalamus can lead to insensitivity toward the leptin signal, thus implying that this protein may be involved in the regulation of this pathway. IRS-2 may therefore be an integral effector molecule by ensuring the appropriate response towards the leptin protein (Burks *et al.*, 2000).

Resistance to leptin via its receptor is a major causative agent of obesity and various disorders such as DM and liver disease. The increased fatty acid level ensuing from the aforementioned resistance due to increased lipolysis, stimulates the TF PPARγ, which induces the expression of proteins such as acetyl-CoA carboxylase (ACC), fatty acid synthetase (FAS) and serine-palmitoyl transferase (see Figure 2.3). The latter enzyme is responsible for the production of ceramide from the condensation of palmitoyl-CoA and

serine. Ceramide induces the expression of nitric oxide synthase, which is in turn a powerful inducer of apoptosis (Unger *et al.*, 1999).

Figure 2.3: Diagrammatic representation of the leptin induced pathway in normal and leptin resistant nonadipocytes



OB-R = leptin receptor; P-STAT-3 = phosphorylated signal transducers and activators of transcription 3; PPAR α and γ = peroxisome proliferator activated receptor α and γ; PPRE = PPAR response element; ACO = acyl CoA oxidase; CPT-1 = carnitine palmitoyl transferase-1; ACC = acetyl CoA carboxylase; FAS = fatty acid synthetase; FA = fatty acid; Adapted from Unger *et al.* (1999).

The islets of Langerhans, hepatocytes, skeletal and cardiac myocytes are major sites of leptin action. Resistance of the receptor to leptin results in apoptosis of these cells, which in turn results in a decrease of insulin production. As the obese person requires higher levels of insulin, this decrease in production results in hyperglycaemia within the patient and eventually causes T2D. Loss of the hepatocytes results in hepatic steatosis or fatty liver while loss of ageing myocytes results in muscle wasting in aged, obese patients (Unger *et al.*, 1999).

Women present with a 40% higher level of leptin when compared to men of the same percent body mass and fat mass (Saad *et al.*, 1997). The effect is not due to the differences in sex hormones, as this effect was still present in postmenopausal women. It was hypothesised that it may be due to altered expression levels of leptin based upon the differential distribution of adipocytes between the sexes, or alternatively it may be due to varied hypothalamic regulation of leptin production. Further investigation is, however, required before this can be stated (Saad *et al.*, 1997).

Further evidence for this effect has been detected in the cord blood of newborns. It was determined that leptin levels were significantly higher in females as well as being positively correlated with birth weight (Matsuda *et al.*, 1997). This was an important correlation, as newborns cannot control their energy uptake by the amount of food they ingest. This is one of the proposed mechanisms by which leptin causes its lipostatic effects i.e. by inducing hypophagia (Pelleymounter *et al.*, 1995). Therefore leptin may preferentially act through the increased catabolism of lipids.

In a similar investigation leptin levels were significantly associated with birth weight, cord blood insulin concentration and placental weight (Koistinen *et al.*, 1997). This indicates that sensitivity to leptin may have a major effect on the intrauterine environment. If the "thrifty phenotype" hypothesis (as discussed in Section 2.1.4.2.1) is taken into consideration, i.e. the fact that there are specific variants which affect T2D susceptibility via induction of a deficient intrauterine environment, leptin and its receptor may be excellent candidates for possible sites of this effect. The role of this protein in T2D disease risk is diagrammatically presented in Section 6.1.

2.1.2.2 Tumour necrosis factor alpha (TNFα)

Tumour necrosis factor α (TNFα) affects weight control via the regulation of adipocyte expression (Hotamisligil *et al.*, 1995). It is a modulator of gene expression in adipocytes and undergoes increased expression in the obese (Hotamisligil and Spiegelman, 1994). Higher levels of expression result in hyperinsulinaemia by the induction of resistance towards insulin. Resistance is due to the Ser phosphorylation of IRS-1, which subsequently decreases the tyrosine kinase activity of the insulin receptor. TNFα also induces the expression of leptin therefore decreasing insulin sensitivity.

2.1.2.3 Adiponectin (APM1)

Dysregulation of APM1 has conclusively been associated with T2D. In an investigation by Spranger *et al.* (2003) it was determined that low plasma levels of APM1 were significantly associated with increased T2D risk whereas an increase in this protein was alternatively associated with an increased protection against developing one of these disorders. Furthermore investigations of the plasma levels of this protein have resulted in the determination that there are much lower levels in obese individuals. APM1 levels have been found to be positively correlated with levels of HDL-C (Zietz *et al.*, 2003). This is an important association as low levels of HDL-C represent an independent cardiovascular risk

factor. As previously stated diabetics often have low levels of APM1 thus increasing the cardiovascular risk of a patient via the associated decreased levels of HDL-C (Zietz *et al.*, 2003)

The APM1 protein itself has been determined to reverse decreased insulin sensitivity brought about by obesity (Yamauchi *et al.*, 2001). Including this glucose sensitising effect the protein is able to increase fatty-acid oxidising activities, both of these functions resulting from the phosphorylation of 5'-adenosine monophosphate activated protein kinase (AMPK). The activation of this protein leads to the phosphorylation of the β-isoform of acetyl coenzyme A carboxylase (ACC-β), which in turn results in the inactivation of ACC. This causes the activation of fatty acid β-oxidation thus resulting in triglyceride release in myo- and hepatocytes. This process is presented in an overview in Section 6.1.

As discussed previously this pathway results in the increased expression of enzymes such as ACO and CPT-1 thus resulting in an increase in fatty acid oxidation. This prevents the accumulation of fatty acids in the form of triglycerides within non-adipocytes such as the β cells (Unger *et al.*, 1999).

AMPK activation causes increased glycogen phosphorylase activity, which preferentially increases the activity of glycolysis thus inducing the down regulation of gluconeogenesis, when exposed to hepatocytes. Thus APM1 exerts its antidiabetic effects by decreasing glucose production in the liver and increasing fatty acid oxidation in muscle cells, thus preventing triglyceride accumulation (Yamauchi *et al.*, 2002). Therefore the association of hypoadiponectinaemia with insulin resistance and hyperinsulinaemia described by Weyer *et al.* (2001) must be caused by the abolition of the aforementioned pathway. Any alteration on a genetic level, which affects the production of this protein, may thus result in impaired insulin action and hence result in T2D. The investigation by Weyer *et al.* in 2001 was the first direct association between the level of APM1 and impaired glucose control. Previously it was hypothesised that the aberrant insulin levels were secondary to the increased obesity, which was induced by the decrease in APM1 levels.

APM1 has also been determined to control the expression of various adhesion molecules such as intracellular adhesion molecule-1 (ICAM-1) by preventing the expression signal induced by TNFα via the nuclear factor kappa (κ) B (NFκB) induced pathway. APM1 is

able to bind to human aortic endothelial cells which in turn results in an increase in the intracellular levels of cAMP levels. The increased production of these molecules prevents the phosphorylation and degradation of the inhibitor of NFκB (IκB) thus causing loss of the TNFα induced signal (Ouchi *et al.*, 2000). This could explain the link between APM1 and CAD. If APM1 levels are low there is an increase in the adhesion of monocytes as well as the expression of the various adhesion molecules. Upon investigation of patients suffering from CAD complications it was determined that the mean plasma APM1 level was significantly lower than in non-affected individuals (Ouchi *et al.*, 1999). Thus APM1 shortage is a possible risk factor in CAD due to it negatively affecting the inhibition of the NFκB pathway.

An important negative correlation has been detected between APM1 and leptin in obese patients in that a decrease in APM1 results in an increase in leptin (Matsubara *et al.*, 2002). This is important as both these proteins have similar functions in enhancing the breakdown of fatty acids (Yamauchi *et al.*, 2002; Unger *et al.*, 1999). It is the view of the author that this negative correlation may be necessary to ensure homeostasis of fatty acid oxidation. This raises a dilemma when APM1 levels are decreased, and the leptin receptors are non-functional, there is no prevention of triglyceride accumulation within nonadipocytes. It will be necessary to screen patients with high adiposity to determine if this interaction is significant in disease pathogenesis.

2.1.3 Clinical features of diabetes mellitus

DM presents with an array of characteristic symptoms. These include thirst, polyuria, blurring of vision, fluctuations in weight, ketoacidosis and a non-ketotic hyperosmotic state (WHO Consortium, 1999). These disorders are, however, often undiagnosed, thus allowing the hyperglycaemia to cause various long term pathological and functional changes such as progressive development of the complications of retinopathy resulting in possible blindness, nephropathy which may result in renal failure, neuropathy, foot ulcers, amputation, various autonomic dysfunctions as well as increased risk towards cardiovascular as well as peripheral vascular and cerebrovascular complications (WHO Consortium, 1999).

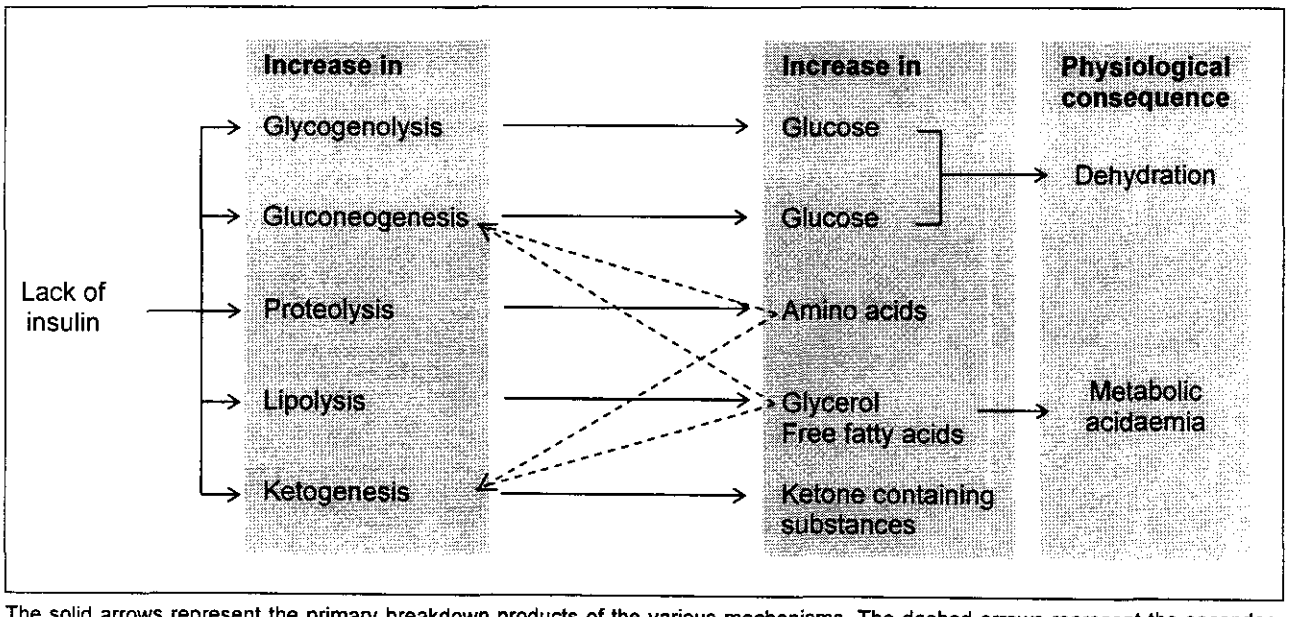
In DM the chronic hyperglycaemia generally results from the non-inhibition of gluconeogenesis due to the absence of or resistance to insulin, thus causing osmotic diuresis. This causes dehydration, which as stated previously, is one of the primary

symptoms of these disorders. Higher levels of proteolysis within muscle produce greater levels of glucogenic amino acids, increasing the concentration of plasma glucose, which in turn aggravates the diuresis (Sonksen and Sonksen, 2000).

Lipolysis is responsible for an increase in the levels of glycerol and FFA. An increase in the former causes hypertonicity in free flowing plasma. The FFAs as well as the ketogenic amino acids are highly insoluble and are therefore transported to the liver. Enzymes within this organ form ester bonds within the molecules to increase solubility, but in doing so the level of ketoacids, e.g. 3-hydroxybutyric acid and acetoacetic acid, increases. Due to the preferential use of ketoacids for energy production as opposed to glucose, the concentration of these molecules increases in the cell. The hydrogen ion (pH) buffering systems present in the circulatory system are not able to maintain physiological pH, culminating in metabolic acidosis. This as well as the osmolar diuresis may induce coma and possibly result in death if not treated (Sonksen and Sonksen, 2000).

Figure 2.4 represents the physiological consequences of a lack of insulin to a patient with DM. The major consequence is the increase in alternate forms of energy production. The end products of various processes undergo secondary breakdown by the primary mechanisms thus increasing the overall effect that the absence of, or a decreased sensitivity towards insulin, has on the patient.

Figure 2.4: Diagrammatic representation of the biochemical consequences of insulin deficiency



The solid arrows represent the primary breakdown products of the various mechanisms. The dashed arrows represent the secondary breakdown products. Adapted from Sonksen and Sonksen (2000).

T2D, CAD and hypertension are associated with the hyperinsulinaemia caused by the resistance towards this hormone (Reavan, 1988). These disorders are due to the effect that insulin has on the FFA concentration. Insulin increases the preferential use of glucose by the body (Brown *et al.*, 1978) therefore resistance towards this hormone increases the level of FFA, specifically very low density lipoprotein (VLDL) in the blood (Olefsky *et al.*, 1974). This is due to lipolysis being the alternative form of energy production when cellular concentrations of insulin are low which, in the case of an affected individual, is caused due to the cellular resistance towards the hormone (Thomas *et al.*, 1979).

Increased blood pressure results due to the direct relationship between the ensuing hypertension and the hyperinsulinaemia (Modan *et al.*, 1985). Increased hypertension is in turn an aggravating factor for CAD. A higher concentration of plasma triglycerides and the decreased concentration of the HDL-C caused by increased insulin secretion are also major aetiological agents (Reavan *et al.*, 1967; Swislocki *et al.*, 1987). This results in the formation of atherosclerotic plaques in the coronary artery thus culminating in its dysfunction (Ducimetiere *et al.*, 1980).

Another cause of hypertension, and eventually CAD, present in the diabetic patient is due to the induction of atherosclerosis via increased levels of glycoxidation of low density lipoprotein (glyc-ox LDL), which is due to the heightened glucose levels (Napoli *et al.*, 2002). This molecule represses the expression of the enzyme nitric oxide synthase III (NOSIII) via a cyclic guanosine monophosphate (cGMP)-dependent pathway. It has been determined that glyc-ox LDL causes this repression via decreasing the binding affinity of the transcription factors specific for the sterol responsive element within the 5' flanking region of NOSIII. This reduction in transcription results in decreased levels of nitric oxide, which is an important repressor of vascular smooth muscle cell proliferation (Napoli *et al.*, 2002). It has been hypothesised by the authors that its absence induces proliferation of these cells therefore decreasing the vasodilatory ability of the arterial tissue and inducing hypertension.

2.1.4 Diagnosis of diabetes mellitus

Numerous conditions may result in the major symptom of hyperglycaemia therefore, although this phenotype is easily recognisable, the diagnosis of DM as the cause is complex. For this reason a single hyperglycaemic reading should never be used as a definitive indication of diabetes in an asymptomatic individual. The diagnostic criteria of

these disorders as well as the various methods by which these values are determined are discussed below (WHO consortium, 1999).

The initial diagnosis is warranted due to the patients presenting with certain definitive characteristics such as increased thirst and urine volume, recurrent infections, weight deviations, drowsiness, coma or high urinary glucose levels. In symptomatic individuals the diagnosis can be confirmed upon a single determination of high blood glucose following a period of fasting. In an asymptomatic individual it is necessary to determine a second reading following either a period of fasting or an oral glucose tolerance test (OGTT). The threshold values for diagnosis as defined by the WHO consortium are indicated in Table 2.2.

Table 2.2: Diagnostic criteria of the hyperglycaemic state

Hyperglycaemic state	Fasting plasma glucose	Whole blood glucose
impaired fasting glycaemia	> 6.1 mmol.l^{-1} but < 7.0 mmol.l^{-1}	> 5.6 mmol.l^{-1} but < 6.1 mmol.l^{-1}
impaired glucose tolerance	< 7.0 mmol.l^{-1}	< 6.1 mmol.l^{-1}
diabetes mellitus	> 7.0 mmol.l^{-1}	> 6.1 mmol.l^{-1}

% = percent; < = less than; > = greater than. Adapted from the WHO Consortium (1999).

The criteria stipulated in Table 2.2 are utilised for the determination of the glycaemic state of an individual. This is important because individuals with impaired fasting hyperglycaemia or glucose tolerance may deteriorate to diabetes mellitus. Another method by which these disorders may be diagnosed is in terms of the aetiology of the disease. This has arisen from an improved understanding of the various causes responsible for diabetes mellitus. This method allows for the classification of this group of disorders into type 1 diabetes mellitus (T1D), T2D and a rather large encompassing sub class, referred to as "other specific types of diabetes".

2.1.4.1 Type 1 diabetes mellitus (T1D)

T1D most frequently affects the young and is due to the degradation of the insulin-producing β cells found in the pancreas by autoantibodies (Bottazzo, 1984). It has been determined that specific human leukocyte antigen (HLA) isoforms i.e. HLA-DR4 are associated with this form of DM therefore suggesting a strong genetic aetiology (Congia *et al.*, 1998). However, in apparent disagreement with this, is the fact that upon investigation of concordance of these disorders in monozygotic twins, greater levels of discordance were determined to be present than would be expected if T1D were a

monogenic disorder (Barnett *et al.*, 1981). This disparity may be due to T1D pathogenesis being determined environmentally either via viral infection, stress or trauma.

2.1.4.2 Type 2 diabetes mellitus (T2D)

T2D is a group of disorders caused by a resistance towards insulin, affecting circa (ca.) 5% of the general population (Polonsky *et al.*, 1996). The major sites of this resistance are the peripheral tissues such as the muscles in the extremities (DeFronzo *et al.*, 1985). Resistance to insulin may be due to both genetic abnormalities as well as metabolic defects e.g. obesity.

It has been determined that in obese patients the level of plasma insulin and glucose, produce an oscillating graph when plotted against time following nutrient intake after a period of fasting, similar to that of normal weight patients discussed previously (Hansen *et al.*, 1982). However, obese patients have significantly higher concentrations of glucose and insulin when compared to that of normal weight individuals (Bergman *et al.*, 1981). This has been determined to be due to the resistance toward insulin by the body resulting in an increased level of this hormone in the plasma of the affected patient (Dohm *et al.*, 1988). As the rate of glucose uptake is decreased by this insensitivity, the β cells in turn must undergo hypersecretion so as to maintain the correct balance in the body. This hypersecretion is the major cause of the increased fasting plasma insulin levels detected in individuals affected by T2D (Faber and Damsgaard, 1984). The hypersecretion eventually results in the loss of these cells, as they are unable to maintain such high levels of production for long periods (Kahn *et al.*, 1993).

It was further determined that with respect to the mean insulin level, the increase in the amplitude of insulin secretion in response to a glucose challenge in obese patients, was less than that of normal weight individuals (Hansen *et al.*, 1982). It is more likely that this is due to an increase in the basal insulin release caused by the hypersecretion of insulin brought about by cellular resistance to the hormone, than a postprandial increase in the release of insulin. Alternatively stated the release of insulin is increased under normal circumstances, thus there is no significant increase in hormone release in response to nutrient intake.

Insulin exacts its effects by increasing glucose transport into the cell and a simultaneous decrease in hepatic glucose production (DeFronzo *et al.*, 1989). In an insulin-resistant

individual, hyperglycaemia occurs primarily due to the decreased uptake of glucose. In adipocytes it has been determined that this is partly due to decreased mRNA and protein levels of glucose transporters, while in skeletal muscle this is due to impaired transporter export to the plasma membrane (Garvey *et al.*, 1992). Hepatic glucose production is also not curbed, therefore aggravating the increased plasma glucose concentration. Although the hepatic glucose production is not necessarily increased, its normoglycaemic pattern in a patient with extremely high blood glucose levels is aberrant.

2.1.4.2.1 The "thrifty phenotype" hypothesis

An important hypothesis with reference to T2D is that of the "thrifty phenotype" (Hales and Barker, 1992). Data collected by these authors presented an increasing negative correlation between birth weights and T2D. It has been hypothesised that due to an insufficient intrauterine environment, foetal malnutrition ensues. There is thus induction of the preferential utilisation of nutrients by the brain at the expense of various other "non-essential" organs, such as the pancreas. Improper development of the islets of Langerhans ensues, thus leading to impaired glucose tolerance and insulin resistance throughout the various tissues. The malnutritive state is maintained during foetal development and postnatally as long as the individual remains undernourished. If the individual is exposed to a normal or excessive diet, the impaired glucose tolerance becomes exacerbated and leads to full blown T2D. As monozygotic twins share the same intrauterine environment this hypothesis also provides an explanation for the high levels of concordance, which have been determined with respect to T2D (Barnett *et al.*, 1981).

The hypothesis, however, does not negate the involvement of genetic susceptibility and *vice versa*. Although it has been determined that individuals with low birth weights were more inclined to develop T2D, it is not authoritative. The opposite of this argument i.e. that individuals with normal to high birth weights develop T2D also holds (Hales *et al.*, 1991; Barker *et al.*, 1990). It is therefore apparent that birth weight is not the only determining factor in disease susceptibility and that understanding the heterogeneity of these disorders is much more complex than previously hypothesised.

2.1.4.3 Other types of diabetes

Diabetes mellitus is an extremely complex group of disorders and thus there are various forms which do not conform to the criteria of T1D or T2D. Thus these forms are collectively

termed "other types of diabetes". These include disorders due to genetic defects in β cell function as well as genetic defects in insulin action.

2.1.4.3.1 Genetic defects of beta cell function

The role of the insulin secretion in the pathogenesis of these diseases is becoming more apparent (Tattersall and Fajans, 1975). The more severe and earlier onset forms of diabetes mellitus are due to an inability to produce this hormone correctly. Thus investigations into the possible candidate genes involved in the correct functioning of β cells has resulted in the elucidation of the genetic basis of the following forms of diabetes mellitus.

2.1.4.3.1.1 Maturity onset diabetes of the young (MODY)

A specific form of T2D, known as maturity onset diabetes of the young (MODY), affects children and adolescents, contrary to the normal pathogenesis followed by the various other forms of T2D. It is inherited in an autosomal dominant mode (Tattersall and Fajans, 1975).

MODY locus 1 (MODY1) is due to alterations within the hepatocyte nuclear factor 4 alpha (HNF4α), gene as determined by the investigation of a German pedigree (Yamagata *et al.*, 1996a). It was determined from the investigation of Yamagata *et al.* (1996a) that a specific alteration termed glutamic acid (Glu) 268 stop codon (X) cosegregated with the MODY1 phenotype and was absent from 108 control individuals (Yamagata *et al.*, 1996a). Specific alterations within the HNF4α gene as presented in Table 2.3, were investigated in order to determine the possible mechanism by which these mutations caused MODY1 (Lausen *et al.*, 2000).

Table 2.3: Causative mutations within the hepatocyte nuclear factor 4α gene

Mutation type	Location	Mutation type	Location
Nonsense	Arg154X	Missense	Asp276Glu
Nonsense	Glu268X	Missense	Arg127Trp
Missense	Val255Met	---	---

Arg = arginine; X = stop codon; Glu = glutamic acid; Val = valine; Met = methionine; Asp = aspartic acid; Trp = tryptophan.

Arg154X retained its DNA binding activity but due to its truncation lost its ability to transactivate the reporter vector. The Glu268X mutation also was not able to transactivate

the reporter vector but in this case it failed to bind the HNF4α binding sites upon nuclear shift assay (Lausen *et al.*, 2000). Previous analysis of this alteration indicated that it resulted in the complete ablation of heterodimerisation and DNA binding ability (Stoffel and Duncan, 1997). Furthermore there is also a downregulation in the expression of certain genes involved in glucose transport and metabolism such as glucose transporter 2 (GLUT2), glyceraldehyde-3-phosphate dehydrogenase and liver pyruvate kinase. It was hypothesised that this causes a defect in the glucose sensing mechanism within the β cell (Stoffel and Duncan, 1997).

The missense mutations on the other hand presented with at least some transactivation ability. Arg127Trp presented with a 50% decrease in its ability to transactivate the reporter vector whereas the other two only presented with decreased transactivation at low levels of the expression vector. This could indicate that the latter two alterations prevent the effects that subtle expression of the HNF4α gene may exhibit (Lausen *et al.*, 2000).

A second means by which genetic alterations could affect the pathogenesis of MODY was uncovered upon investigation of a mutation present at a position upstream of the hepatocyte nuclear factor 1 alpha (HNF1α) gene. It was determined that this alteration i.e. a single nucleotide change at nucleotide position 58, resulted in an approximately 30-fold decrease in the ability of wild-type HNF4α to activate the HNF1α promoter. This provides evidence for the fact that alterations within the hepatocyte nuclear factor signalling pathway are intimately involved in MODY progression (Lausen *et al.*, 2000).

The origins of maturity onset diabetes of the young locus 3 (MODY3) were determined by the screening of all genes within the region of chromosome 12q24.2, which had previously been linked to this disorder, in various MODY3 pedigrees (Mahtani *et al.*, 1996). Various alterations within the HNF1α gene were determined to co-segregate with the MODY3 phenotype within the pedigrees as opposed to the various other genes screened within the locus of interest. It was therefore hypothesised that the linkage to the MODY3 locus was due to alterations within the HNF1α gene (Yamagata *et al.*, 1996b). Analysis of the gene encoding HNF1α in French Caucasian MODY3 families resulted in the identification of ten mutations which co-segregate with the disease in specific families (Vaxillaire *et al.*, 1997). Of these ten only a single alteration had been detected in a prior investigation (Yamagata *et al.*, 1996b). The remaining nine mutations determined in the investigation by Vaxillaire *et al.* (1997) were each only present in a single family. These variants may therefore not

be important in susceptibility at the population level. Upon functional analyses, however it was determined that mutations within the MODY3 gene disrupt dimerisation of this protein with itself thus preventing correct expression. These mutations cause disturbances within the four-helix bundle of the dimerisation domain thus resulting in the loss of HNF1α activity (Hua *et al.*, 2000).

In an investigation to determine whether HNF1α was unique, or if it was a member of a family of homeodomain containing transcription factors, a second closely related TF known as hepatocyte nuclear factor 1β (HNF1β) or maturity onset diabetes of the young locus 5 (MODY5) defined by Froguel and Velho (1999), was discovered. This protein was found to form dimers with HNF-1α and it was further able to recognise similar binding sites (Bach *et al.*, 1991).

The phenotype of deletions or insertions within this gene seems to be different from the standard MODY phenotype (Lindner *et al.*, 1999). Screening of Norwegian MODY families resulted in the determination of a deletion of the region encoding amino acid residue 137 to 161. It resulted in the deletion of a region involved in DNA binding specificity. Individuals within this family harbouring this alteration present with diabetes mellitus as well as non-diabetic nephropathy. Certain female members of the family harbouring the alteration also had genital malformations. It was hypothesised that due to the fact that the transactivation domain was still present within the mutated HNF1β it may be activating a different group of genes thus causing the alteration in genital formation. Further studies will, however, be required before this can be proven with any certainty (Lindner *et al.*, 1999).

Evidence for the involvement of the alteration within HNF1β in disease phenotype i.e. the inclusion of renal disease to the general MODY phenotype, was further determined upon investigation of a group of 40 subjects with early onset diabetes mellitus. An insertion mutation, detected in a single family, of a GG nucleotide at the codon encoding amino acid residue 263 caused a frameshift mutation and was therefore named A263GG. The clinical phenotype for individuals harbouring this mutation included diabetes mellitus as well as some form of renal dysfunction. Genital malformations were not investigated, however, it is unlikely that the authors would not have noted this if it had indeed been present in the various affected individuals. It therefore garners further evidence for the hypothesis that

the transactivation domain can still be active following deletion, as in this example the transactivation domain was ablated due to the frameshift (Nishigori *et al.*, 1998).

Analysis of mice deficient for neurogenin differentiation factor 1 (NEUROD1) also termed maturity onset diabetes of the young locus 6 (MODY6) resulted in the finding that this protein is necessary for pancreatic islet morphogenesis thus its absence causes ablation of the insulin secretory ability of the β cell (Naya *et al.*, 1997). In the murine model of insulin promoter factor 1 (IPF1) it was determined that ablation of this protein via a gene knockout strategy similarly resulted in the complete absence of pancreas formation (Jonsson *et al.*, 1994) thus indicating that this protein is integral to the correct development of this organ. IPF-1 is furthermore a direct transactivator of insulin expression within pancreatic β cells (Ohlsson *et al.*, 1993) thus indicating a direct mechanism by which mutations within IPF-1 or maturity onset diabetes of the young locus 4 (MODY4) as defined by Froguel and Velho (1999) could result in decreased insulin secretion. The expression of the protein itself in pancreatic cells is controlled by the binding of TF hepatocyte nuclear factor 3 β (HNF3 β) as well as NEUROD1 (Sharma *et al.*, 1997). It is thus apparent that most of the genes involved in the various forms of MODY encode transcription factors involved in the correct differentiation and functioning of β cells.

The interaction of these various transcription factors alludes to the reason why alterations within the various encoding genes result in T2D. It is a fact that most of these factors act via a similar pathway. HNF4α has an alternative promoter region at ca. 45 kb upstream of the initiation site (Thomas *et al.*, 2001), which is active mainly within β cells. Site directed mutagenesis of a putative HNF1α and β binding site within this promoter element resulted in a decrease in transactivation following transfection analysis thus indicating that these TF are integral in the correct activation of this element. It was further determined upon mutation analysis that a T-146C alteration within this region, which cosegregated with MODY, resulted in a three fold decrease in protein expression. Utilising double transfection assays, as well as nuclear shift assays it was recognised that this alteration affected the binding of IPF-1. Thus all three of these transcription factors are integral in the activation of HNF4α expression within the β cell (Thomas *et al.*, 2001).

The glucokinase gene or maturity onset diabetes of the young locus 2 (MODY2) was located on chromosome 7p (Froguel *et al.*, 1992). The protein glucokinase catalyses the production of glucose-6-phosphate from the substrate glucose during the glycolytic

pathway via ATP mediated phosphorylation. The aforementioned protein is only expressed in the liver and pancreatic β cells (Matschinsky, 1990). It is assumed that this protein controls glucose uptake in the liver and is involved in determination of the presence of glucose and therefore regulation of insulin in the pancreas (Magnuson, 1990). Mutations within glucokinase have been determined to generally cause down regulation of this protein due to for example incorrect ribonucleic acid (RNA) splicing, nonsense and missense mutations which have all been associated with the MODY phenotype (Froguel *et al.*, 1993).

2.1.4.3.1.2 Maternally inherited forms of type 2 diabetes mellitus (T2D)

T2D has a higher proportion of maternally inherited forms, hence it was necessary to elucidate the basis of this skewed ratio (Alcolado and Alcolado, 1991; De Silva *et al.*, 2002). The first studies undertaken focussed on the mitochondrial (mt) genome, due to its maternal mode of inheritance (Giles *et al.*, 1980). It has been determined that various genetic alterations within the mitochondrial genome are associated with T2D, from large deletions (Ballinger *et al.*, 1992) to single nucleotide mutations (van den Ouweland *et al.*, 1994).

As mitochondrial disorders are generally associated with mutations in the tRNAs, it was reasoned that these genes would be most likely affected in mt T2D (Thomas *et al.*, 1996). Diabetes mellitus has been associated with the A3243G mutation within the mt tRNA recognising the UUR codon to encode the amino acid leucine or $tRNA^{Leu\ (UUR)}$ gene, which in turn is associated with MELAS or mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (van den Ouweland *et al.*, 1994).

The actual mechanism involved in the pathology of this mutation is not fully understood. It has been determined that the mutation causes a modification in the leucine incorporation of mitochondrially encoded polypeptides (Flierl *et al.*, 1997). This causes biochemical and structural changes via the loss of the proton gradient due to the incorrect formation of the complexes involved in mitochondrial respiration. In turn this results in dysfunction of the oxidative phosphorylation (OXPHOS) system (Flierl *et al.*, 1997). Another possible mechanism may be due to the mutation being present in the middle of a tridecamer sequence that is involved in the formation of the 3'-end of the 16 Svedberg units (S) ribosomal unit that is located next to the $tRNA^{Leu\ (UUR)}$ gene (Christianson and Clayton, 1988). This results in the loss of termination of transcription of the 16S ribosomal RNA subunit (16S rRNA) as described by Hess *et al.* (1991). The clinical phenotype of this

mutation is believed to be due to the inability to manufacture the exact type and amount of the 16S rRNA, which affects the correct translation of the other gene products due to incorrect ribosome formation (Hess *et al.*, 1991).

This mutation and its effect on β cells of the pancreas have been well investigated. It was demonstrated that the cells contained decreased levels of cytochrome c oxidase (COX), which is a protein partly encoded by the mt genome. It is hypothesised that this induces increased levels of hydroxyl radicals that result in dysfunction of the cell. This in turn explains the loss of function of these pancreatic cells when insulin is hypersecreted in response to the insulin insensitivity (Kobayashi *et al.*, 1997).

This mutation was determined to be present in 1% of diabetic patients in various non-African populations (Vionnet *et al.*, 1993). Few investigations have been undertaken to determine the importance of this mutation in the African population and, to date, it has yet to be detected in the African diabetic population (Olckers *et al.*, 2001; Towers *et al.*, 2002).

2.1.4.3.2 Genetic defects in insulin action

In an investigation by Freidenberg *et al.* (1987) it was determined that the activity of the insulin receptor was down regulated in patients with T2D. There are five major classes of mutations that result in this decreased activity of the insulin receptor gene. The first group includes mutations that result in impaired receptor biosynthesis (Taylor, 1992). These include mutations that affect the transcription of the gene, mutations that cause premature termination of the protein (Kadowaki *et al.*, 1990a), as well as mutations in *cis*-acting elements that are involved in the regulation of transcription of this gene (Kadowaki *et al.*, 1990b). The second group consists of mutations that result in the defective transport of the receptor to the plasma lemma (Kadowaki *et al.*, 1991). This process is altered by mutations affecting the post-translational processing of this protein e.g. incorrect assembly, incorrect cleavage of subunits, lack of O-glycosylation, acylation and terminal processing. The third class of mutations result in increased degradation of the protein. The fourth class affects the ability of insulin to bind to its receptor (Kadowaki *et al.*, 1990c) and the fifth class affects the tyrosine kinase function of the protein (Moller *et al.*, 1991). All of these mutations result in decreased sensitivity towards the hormone insulin and can cause T2D but to date have been mostly associated with such disorders as type A insulin resistance (polycystic ovary syndrome) and acanthosis nigricans (Moller and Flier, 1988).

2.1.5 Treatment of diabetes mellitus

The various major treatment strategies to counter the symptoms of diabetes mellitus are briefly discussed in this section. It shall include the role of diet and physical activity (Eriksson and Lindgärde, 1991), followed by an overview of the various pharmaceutical treatment regimens (Nolan *et al.*, 1994). Although treatment of diabetes mellitus is quite efficient, all of these strategies are long-term solutions. An understanding of the molecular basis of these disorders will allow the development of pre-symptomatic and intervention treatment modalities, that will prevent serious future complications.

One of the simplest methods of treatment, nutritional intervention (often including increased activity) is also the most effective tool to regulate the pathogenesis of T2D (Eriksson and Lindgärde, 1991). This form of therapy is generally utilised in tandem with a pharmaceutical strategy. It reduces fasting plasma glucose and increases oxygen uptake, which results in increased glucose regulation. This reduces the risk of individuals with impaired glucose tolerance (IGT) deteriorating to a T2D phenotype. Individuals undergoing such treatment are also more likely to present with a decrease in body weight, which is associated with decreased triglyceride levels as well as decreased hypertension risk. The Therapeutic Lifestyle Changes (TLC) diet as depicted in Table 2.4 is often implemented in the treatment of the metabolic syndrome as a whole. Diet and physical exercise are therefore important therapeutic tools in the treatment of not only T2D but also the various associated complications (Eriksson and Lindgärde, 1991).

Table 2.4: Nutrient Composition of the Therapeutic Lifestyle Changes (TLC) Diet

Nutrient	Recommended Intake
Saturated fat	> 7% of total calories
Polyunsaturated fat	Up to 10% of total calories
Monounsaturated fat	Up to 20% of total calories
Total fat	25-35% of total calories
Carbohydrate	50-60% of total calories
Fibre	20-30 grams per day
Protein	Approximately 15% of total calories
Cholesterol	< 200 milligrams per day
Total calories	Balance energy intake and expenditure to maintain desirable body weight and prevent weight gain

% = percent; > = greater than; < = less than; Adapted from the Expert Panel on Detection, Evaluation and Treatment of High Blood Cholesterol in Adults (2001).

Another treatment of T2D is via the use of oral hypoglycaemic inducing agents. There are two major families of drugs utilised in the treatment of these disorders. Sulphonylureas cause down regulation of glucose by increasing insulin secretion and activity thus these molecules are capable of inducing a hypoglycaemic state. To overcome this, fast acting sulphonylureas are generally utilised (Knowler *et al.,* 1995).

The second category of pharmaceuticals is the biguanides, which prevent absorption of glucose by the intestine as well as increasing insulin action (Hother-Nielsen *et al.*, 1989). Biguanides and metformin in particular, are excellent therapeutic compounds to be utilised in the treatment of obese patients affected by T2D. It has been determined that this compound causes increased sensitivity to insulin within the peripheral systems but does not affect hepatic sensitivity. It thus increases glucose uptake, but there is no concomitant decrease in hepatic glucose production (Hother-Nielsen *et al.*, 1989).

The pharmaceuticals that act against diabetes directly are mainly thiazolidinediones and glucosidase inhibitors. The former increases the action of insulin (Nolan *et al.*, 1994) while the glucosidase inhibitors delay absorption of glucose and decrease FFA and glucose concentrations.

Insulin treatment is the final means of treatment for type 1, and severe type 2 diabetics, which, have become completely insulin-dependent and are no longer able to produce this hormone autonomously (Scarlett *et al.*, 1982). It has been determined that the infusion of extraneous insulin results in a significant decrease in plasma glucose levels and urinary glucose excretion. This is due to the fact that upon infusion of insulin, the relative concentration is extremely high thus overcoming the resistance that led to the onset of the disorder in the latter case. As this is an exogenous source, the β cells are not affected thus preventing further depletion (Scarlett *et al.*, 1982).

CHAPTER THREE

THE GENETIC ASPECTS OF TYPE TWO DIABETES MELLITUS

Due to the extreme phenotypic variability of these disorders many researchers have postulated that it was generally due to environmental causes such as obesity (Bergman *et al.*, 1981) or low birth weight (Hales *et al.*, 1991). It has however been determined that T2D presents with high levels of concordance between monozygotic twins after a certain age (Barnett *et al.,* 1981). Another investigation between affected sib-pairs determined that there is a great deal of similarity in the intra familial age of disease onset as compared to interfamilial variation (Lev-Ran *et al.*, 2000). Thereafter, a certain threshold is attained at a specific age and the disorder ensues.

Elucidation of the genetic component involved has resulted in the determination of various genetic susceptibility loci which place an individual at greater risk of contracting one of these disorders (Kadowaki *et al.*, 1990a; Vestergaard *et al.*, 1993). Thus it may be hypothesised that patients with T2D do not necessarily have a monogenic cause but rather various inherited loci, which generate susceptibility towards T2D.

3.1 CANDIDATE GENES FOR TYPE 2 DIABETES MELLITUS SUSCEPTIBILITY

Genetic investigation of complex disorders can be simplified by firstly, screening specific genes known to be involved in the pathways responsible for disease pathogenesis. In Chapter Two the various signalling pathways involved in insulin sensitivity and secretion were discussed. Reference will be made to the associations determined between these candidate genes and T2D. An overview of the most important genes is presented in Section 6.1.

3.1.1 The insulin receptor substrate 1 (IRS-1) gene

Murine Irs-1 was utilised to probe a human male placenta genomic library in order to determine the gene encoding IRS-1, of which five clones presented with positive signals (Stoffel *et al.*, 1993). The gene itself was in turn localised via the polymerase chain reaction (PCR) utilising primers derived from the positive clones, to amplify the gene from

a human-hamster somatic cell hybrid panel. It was by this method that the gene was localised to chromosome 2q. Further fluorescent *in situ* hybridisation (FISH) analyses resulted in the localising of the gene to 2q35-q36.1 (Stoffel *et al.*, 1993). Eventually the IRS-1 gene was localised to chromosome 2q36. This was determined through the application of FISH (Nishiyama *et al.*, 1994).

Determination of genetic alterations within this gene has resulted in numerous candidates such as the glycine (Gly) 818 Arg and Ser892Gly mutations within the Finnish population (Laakso *et al.*, 1994), a Gly723del mutation in a single Italian patient (Esposito *et al.*, 1996) and a Gly972Arg mutation within the Danish (Almind *et al.*, 1993) and Dutch ('t Hart *et al.*, 1999) populations. It has however only been the latter which has been determined to be relevant to disease pathogenesis (Clausen *et al.*, 1995; Baroni *et al.*, 1999; Porzio *et al.*, 1999). This mutation was originally detected in a Danish case-control investigation into possible variants within the IRS-1 gene that were associated with T2D, although the alteration itself was not in association with this disease (Almind *et al.*, 1993).

Functional analysis of a Gly972Arg has resulted in the determination of a possible role in increased T2D susceptibility (Almind *et al.*, 1996). In an investigation of a cell line deficient in IRS-1 and IRS-2 it was determined that transfection with a vector containing this mutant allele resulted in decreased DNA synthesis in comparison to the effect of the wild type. Upon further analysis it was determined that this altered form resulted in a decrease in the activation of the PI3K pathway via the impaired binding of the p85 regulatory subunit thus preventing proper insulin signalling. The cell line utilised did not express IRS-2, which may have had a modulatory effect *in vivo* thus masking the effect of this alteration in patients. The Gly972Arg mutation does however increase susceptibility via this mechanism but it is incapable of causing T2D on its own (Almind *et al.*, 1996).

Investigations into the insulin induced processes which are negatively affected by the Gly972Arg alteration have led to the elucidation of effects at both the glucose transport level as well as glycogen synthesis. Cell lines in which the mutant IRS-1 protein was over expressed presented with decreased GLUT1 and GLUT4 transport to the plasma membrane in comparison to cells over expressing the wild type. This was associated with impaired glucose transport at both the basal and insulin induced levels. Furthermore GSK-3 activity was increased in cells over expressing the variant allele, thus ensuring repression of the glycogen synthase protein. In turn this results in resistance towards the effect of insulin (Hribal *et al.*, 2000).

Possible association of this alteration with other symptoms of the metabolic syndrome was investigated (Baroni *et al.*, 1999). Individuals suffering from CAD were screened for this alteration in comparison to a population-controlled cohort. It was determined that this alteration was significantly associated with increased risk to CAD (p value < 0.001). Furthermore the association of this alteration with T2D was strengthened (p value < 0.01) although this was not the major clinical factor upon which patients were selected. As CAD and T2D are both major symptoms of the metabolic syndrome (Reavan, 1988) it is not surprising that certain patients were affected by both CAD and T2D. Presence of the variant allele was further associated with an approximately 60% increase in triglyceride levels (p value< 0.001). As the major effect of this alteration would seem to be decreased insulin sensitivity, due to impaired signalling, it is reasonable to assume that its association with T2D, CAD and increased triglyceride levels is due to all three of these clinical factors resulting from the metabolic syndrome which is induced by insulin resistance (Reavan, 1988).

An investigation of the Danish population (sample size or $n = 380$) resulted in the determination of an important modulating effect with regard to this alteration (Clausen *et al.*, 1995). The alteration was present in 9% of the individuals screened. If the individual harbouring the alteration was lean it was discovered that the individual presented with similar levels of insulin sensitivity as wild type lean allele carriers. Alternatively if the individual was obese it was calculated via multivariate analysis that both of these factors congruently resulted in a significant 50% increase in insulin resistance (p value $= 0.0008$). This is a simple yet effective example of the effect that environmental factors have on predisposing genetic alleles.

As discussed previously in Section 2.1.1 decreased PI3K signalling would affect both the sensitivity towards insulin as well as insulin secretion. A concomitant investigation exploring the effect that this alteration has on β cells has determined that this is in fact the case as there is a decrease in the insulin content of cells expressing this variant in association with a decrease in p85 binding (Porzio *et al.*, 1999). This alteration within IRS-1 may therefore be more significant in the β cell response to insulin, if such factors as its modulation by obesity are taken into consideration.

Evidence for this hypothesis was provided by a case-control analysis within the Danish T2D population (Almind *et al.*, 1993). No significant association was detected between the

alteration and T2D susceptibility however individuals harbouring the alteration did present with lower levels of insulin and c-peptide but did not present with insulin resistance. The absence of insulin resistance is possibly due to the fact that the overall body mass index (BMI) of individuals harbouring the allele was 29 kilogram.metre^{-2} (kg.m^{-2}), indicating that the individuals are overweight but not obese. It is possible that the Gly972Arg mutation alters correct β cell signalling which does not result in T2D, unless an individual is obese and therefore more likely to be insulin resistant. The effects that this mutation has upon insulin signalling may have a confounding effect upon insulin sensitivity in association with obesity. This resistance in a normal individual would result in increased insulin secretion to ameliorate the effects (Bergman *et al.*, 1981), but in an individual harbouring the mutation there is a decreased sensing capability and thus no response to the resistance.

3.1.2 The insulin receptor substrate 2 (IRS-2) gene

The effect of environmental modifiers on the phenotypic expression of an alteration associated with T2D was made evident in an investigation of an alteration in the IRS-2 gene. Via various association analyses it was determined that the Gly to asparagine (Asn) change at amino acid residue 1057 (Gly1057Asn) is associated with decreased T2D risk in individuals with a BMI < 27 kg.m^{-2}, however in individuals who were overweight, the alteration was associated with an increased risk in both the hetero- and homozygous forms (Mammarella *et al.*, 2000). It was hypothesised that the alteration may have an effect on the ability of various regulatory elements to interact with the IRS-2 protein, therefore preventing its down regulation. This increases the efficiency of the insulin-signalling cascade in normal weight individuals, which in turn increases insulin sensitivity. This effect may be reversed in obese individuals as various other regulatory elements affecting IRS-2 signalling, are triggered via the increased activity of the various adipocytokines brought about by increased fat mass in these individuals (Mammarella *et al.*, 2000).

That this effect was more likely to be due to a defect in the sensitivity towards insulin rather than a deficit in insulin secretion was determined in 2001 by Fritsche *et al.* The insulin secretory ability of β cells in a group of non diabetic individuals (n = 77) were investigated under normal circumstances (i.e. via an OGTT) as well as under extreme stress (i.e. via a hyperglycaemic clamp) in order to detect any latent abnormalities that may be present. It was determined that β cell function was similar for carriers of the mutant allele compared to the non-carriers. It therefore provides evidence, which indicates that the

mechanism of action of this alteration is unlikely to be due to altered insulin secretion (Fritsche *et al.*, 2001). It has been previously discussed that loss of the IRS-2 gene results in deficiencies in both insulin sensitivity and secretion (Withers *et al.*, 1998; Etgen *et al.*, 1999). Therefore with regard to the effects of this alteration, it becomes plausible that different domains of the protein have different roles within the signalling pathway. Therefore if one region is altered by mutation it may only affect certain signalling pathways within the organism.

3.1.3 The mitogen activated protein kinase 8-interacting protein 1 (MAPK8IP1) gene

The importance of mitogen activated protein kinase 8 interacting protein 1 (MAPK8IP1) in the pathogenesis of T2D became evident upon the repression of the translated product via an antisense RNA strategy in an insulinoma β cell line. Both the expression of insulin and GLUT2 were decreased upon expression of the antisense RNA, thus indicating a possible mechanism by which alterations within this gene could result in increased susceptibility (Waeber *et al.*, 2000).

Large-scale sib pair analysis of 149 multiplex T2D families indicated that MAPK8IP1 is not a likely candidate for T2D susceptibility. A single family was however determined to harbour a Ser to asparagine (Asn) alteration, Ser59Asn, that segregated with T2D (maximum logarithm of the odds score (LOD_{max}) = 1.34; p value = 0.005), which upon functional analysis was determined to be less efficient at preventing c-Jun amino terminal kinase from repressing insulin expression. Furthermore, upon co-expression of the mutated MAPK8IP1 with MEKK1 it was determined that the mutated protein was less effective at preventing MEKK1-induced apoptosis (Waeber *et al.*, 2000). This indicates that there are two possible mechanisms by which alterations in this gene can result in impaired insulin secretion and therefore greater T2D susceptibility.

3.1.4 The protein phosphatase 1 regulatory subunit 2 (PPP1R2) gene

In 1976, Huang and Glinsmann discovered two inhibitors of protein phosphatases in rabbits, one of which also seemed to inhibit glycogen synthase phosphatase ability. This protein was termed inhibitor-2. Utilising a yeast two-hybrid system it was possible to determine the human form of the complementary DNA (cDNA) encoding this protein. Upon determination of the protein it was expressed in *Escherichia coliform* (*E. coli*), purified and demonstrated to repress PP1 (Helps *et al.*, 1994).

Utilising this sequence information it was possible for Permana and Mott (1997) to localise the protein phosphatase 1 regulatory subunit 2 (PPP1R2) gene to chromosome 3q29. This was achieved by screening a human foreskin fibroblast genomic library utilising primers specifically designed for the 3'-noncoding region of the PPP1R2 gene. The relevant clones were utilised to develop a FISH assay, which in turn allowed for the identification of the chromosomal locus. The sequencing strategy utilised to elucidate the exons as well as the intronic boundaries was applied to two cohorts consisting of insulin sensitive and insulin resistant Pima Native Americans respectively. No alterations within the coding region were, however, detected (Permana and Mott, 1997). Future studies should include the intronic regions as well as the screening of different populations.

3.1.5 The protein phosphatase 1 regulatory subunit 3A (PPP1R3A) gene

The protein phosphatase 1 regulatory subunit 3A (PPP1R3A) gene which encodes the glycogen bound subunit of PP1 was determined upon screening a human skeletal muscle cDNA library utilising the rabbit cDNA sequence as a probe (Chen *et al.*, 1994). Subsequent to the identification of the human cDNA of the gene, a single strand conformation polymorphism (SSCP) strategy was utilised to screen a Caucasian control and diabetic cohort for alterations. A single patient who concomitantly presented with low levels of nonoxidative glucose metabolism (glycogen synthesis) harboured a C2792A substitution as determined upon sequence analysis of the variant SSCP profile that was detected. This alteration causes an alanine (Ala) 931Glu substitution in the protein. No other variant SSCP profiles were detected for either the patient or control cohort. Furthermore, quantification of mRNA levels of this gene indicated that there was no difference between the subset of control and affected individuals investigated (Chen *et al.*, 1994).

Analysis of a native Canadian population known as the Oji-Cree elucidated the presence of a 5 base pair (bp) insertion/deletion mutation within the mRNA stabilising AU(AT)-rich element of the 3'-untranslated region of the PPP1R3A gene. The homozygous deleted allelic combination was significantly associated with lower levels of glucose at 120 minutes following a glucose challenge in T2D and IGT individuals but was not associated with T2D, IGT or any other clinical factor tested. It is necessary that larger samples be analysed in a future study, as the frequency of homozygotes were low and the obtained p-value was only marginally significant (Hegele *et al.*, 1998).

Few investigations address the interplay between various biological molecules, however the relevance of this type of interaction is of vital importance to biological systems where an interconnected web of metabolic pathways are continuously being synchronised. In an investigation by Savage *et al.* (2002) a family with severe impaired insulin action had various candidate genes screened and mutations were determined in two seemingly unrelated genes, specifically PPARγ and PPP1R3A. Compound heterozygotes for these alterations presented with severe insulin resistance, acanthosis nigricans and variably with T2D, dyslipidaemia or hypertension. Heterozygotes and homozygous wild type individuals presented with normal insulin action, however one heterozygous individual harbouring the mutation within the PPP1R3A gene presented with acanthosis nigricans and severe obesity (Savage *et al.*, 2002).

It has been determined that the mutation within PPARγ causes truncation of the protein, thus preventing it from recognising its specific nucleic acid element. This may result in an increase in the preferential utilisation of fatty acids over glucose as an energy source. Alternatively, the mutation within the PPP1R3A gene causes loss of localisation of the protein to the sarcoplasmic reticulum thus preventing the activation of glycogen synthase. The cumulative effects of these alterations induce the insulin resistant phenotype, leading to the severe hyperinsulinaemia (Savage *et al.*, 2002).

3.1.6 The glycogen synthase (GYS1) gene

Glycogen synthase (GYS1) is another candidate gene associated with T2D (Vestergaard *et al.*, 1993). It is important in the production of glycogen in mammalian skeletal muscle as stated previously in Section 2.1.1 (Dent *et al.*, 1990). There are two pathways by which this protein can cause T2D. Firstly, if there is a mutation that prevents dephosphorylation, the protein will not be able to maintain a high enough level of glycogen synthesis, resulting in an accumulation of glucose. The second possible mechanism may be a mutation in the insulin receptor gene, which results in the cellular kinases and phosphatases not being activated and thereby preventing increased glycogen synthesis and decreased glucose levels (Vaag *et al.*, 1992). This defect is placed in context within Figure 6.2.

The gene encoding this protein was localised to chromosome 19 via screening of a human-hamster somatic cell hybrid panel (Groop *et al.*, 1993). Furthermore an alteration termed T1668C was significantly associated with T2D (p value < 0.001) in the Finnish

population as well as an increased susceptibility toward hypertension in both the diabetic (p value = 0.008) and control (p value = 0.013) cohorts (Groop *et al.*, 1993). Upon investigation of glucose metabolism it was determined that synthesis of glycogen was severely affected within individuals harbouring the variant allele in both the heterozygous and homozygous form. There is however no concomitant decrease in the expression of the protein in skeletal muscle thus indicating that the effect may be at the level of protein function. This is not necessarily due to the alteration itself as it may merely be in linkage disequilibrium with the causative genetic alteration (Groop *et al.*, 1993). In contrast to this investigation a similar analysis of a French T2D population resulted in the determination that the variant allele was lower in patients than in control individuals (Zouali *et al*, 1993). This seeming inconsistency may be due to possible variations in genetic susceptibility between the two populations, albeit that both share similar origins. A more likely explanation is that the individuals investigated in the French population were specified as multiplex families. It is therefore possible that these patients present with MODY as well as T2D. As the pathogenesis within MODY is at the level of insulin secretion (Ohlsson *et al.*, 1993), it is unlikely that this form of the disorder is associated with an alteration which affects insulin sensitivity e.g. the variant allele within the glycogen synthase gene. Further investigation is required before the definitive mechanism can be delineated.

Microsatellite analysis of chromosome 19q in a diabetic and a control cohort of the Pima Native American population resulted in the determination that certain genotypes of the GYS1 specific marker produced were significantly associated with T2D (Majer *et al.*, 1996). Upon analysis of the various alleles it was determined that only three were common in the cohorts investigated namely alleles 5, 6 and 7, with allele sizes of 88, 86 and 84 base pairs respectively. In addition these three alleles were determined to be in association with the disorder upon chi-square (χ^2) analysis. The gene itself however did not harbour any alterations, which were associated with disease risk although it may be that the genetic lesion was present within the promoter region of this gene, which was not investigated. Levels of mRNA were no different between the diabetic and non-diabetic individuals investigated as well as the fractional activity of the enzyme at basal levels. It was however determined that there was a 25% decrease in protein content in diabetic individuals upon western blot analysis. It was hypothesised that this was due to a possible post-translational effect, which alters the epitope so as not to be recognised by the antibody utilised. This effect may be decreased in diabetic individuals (Majer *et al.*, 1996). There was however no investigation into the mRNA, protein and fractional activity of this protein in response to insulin. Thus it cannot be excluded that the 25% decrease in protein

content is actually present within diabetic individuals, but at basal levels there is sufficient enzymatic activity. Upon insulin treatment, the level of protein is too low to attain the activity levels of unaffected individuals and thus hyperglycaemia ensues.

An investigation into monozygotic twins discordant for T2D resulted in identifying that the activity of glycogen synthase in response to insulin is severely decreased within the diabetic twin (Huang *et al.*, 2000). Further analysis of expression indicated that levels of glycogen synthase mRNA were decreased in diabetic individuals in response to insulin stimulation as compared to unaffected individuals. It was therefore hypothesised that the defect in glycogen synthesis was a secondary effect of the lesion causing insulin resistance, instead of an actual genetic defect within the gene itself (Huang *et al.*, 2000). The choice of subjects, i.e. the utilisation of monozygotic twins discordant for the diseased phenotype, would bias this observation towards acquired effects over heritable effects as the phenotype would thus be due to environmental factors instead of genetic susceptibility.

3.1.7 The glycogen synthase kinase-3 alpha (GSK-3α) gene

Human-hamster somatic hybrid cell panels were utilised in order to localise the gene encoding glycogen synthase kinase-3 alpha (GSK-3α) to chromosome 19. The gene was further localised to chromosome 19q13.1-q13.2. Molecular screening of the coding regions of this gene via SSCP analysis of 98 T2D patients and 12 healthy controls resulted in the determination of two silent mutations each in a single individual. It was therefore deemed unlikely that either alteration had a large effect on disease susceptibility. The intronic and promoter regions were not investigated and may harbour alleles, which increase T2D risk (Hansen *et al.*, 1997).

3.1.8 The Obese (ob) gene

The *Obese* (ob) gene, which encodes leptin, has been localised to human chromosome 7q31.3 through FISH analysis. Via a positional cloning approach it was determined that the gene itself was ca. 20 kilobases (kb) in length and consisted of three exons separated by two introns. Further analyses decreased this length to ca. 18 kb, which in turn was determined to encode a 3.5 kb mRNA fragment (Gong *et al.*, 1996). The 5'-untranslated region of the gene harboured a general TATA promoter element as well as binding sites for various *cis*-acting elements e.g. CCAAT enhancer binding protein (C/EBP) binding site (Isse *et al.*, 1995). Transfection assays led to the determination that the human promoter is active in differentiated murine adipocytes (Gong *et al.*, 1996).

3.1.9 The leptin receptor (LEPR) gene

Via the use of human-hamster somatic hybrids it was determined that the gene encoding the leptin receptor was localised to chromosome 1. Scanning a yeast artificial chromosome (YAC) library via PCR analysis resulted in the localisation of this gene to chromosome 1p31 (Winick *et al.*, 1996).

The leptin receptor has been identified to be important in the pathogenesis of T2D and, to date, its role has not been defined. Important mutations in this gene include both nonsense mutations as well as mutations affecting the correct splicing of this protein. Two major mutations under investigation are A861G and G2161C (Chung *et al.*, 1997). The former is located in exon 4 and results in a Glu223Arg which is a neutral amino acid replaced by a positive amino acid, while the latter is responsible for a lysine to an asparagine change which is a replacement of a positive amino acid to a neutral amino acid. The exact mechanism of action whereby disease is caused by these mutations has not been determined but preliminary studies have allowed for the determination that the above mutations are involved in obesity and hence T2D (Chung *et al.*, 1997).

3.1.10 The fatty acid binding protein 2 (FABP2) gene

Insulin sensitivity is highly relevant to T2D pathogenesis (Reavan, 1988). It is thus necessary to identify loci involved in affecting insulin action to elucidate putative susceptibility loci. A significant association was determined upon sib-pair analyses within a Pima Native American cohort (Prochazka *et al.*, 1993). Insulin action at maximal stimulating insulin concentrations was significantly linked (p value = 0.0008) to a genetic marker at the fatty acid binding protein 2 (FABP2) locus (Prochazka *et al.*, 1993). The locus harbouring the FABP2 gene was thus likely to be a candidate for T2D risk.

Molecular investigation of this gene resulted in elucidation of the Ala54Thr alteration within the Pima Native American cohort (Baier *et al.*, 1995). Comparison of the homozygous carriers of the wild type allele and the homozygous and heterozygous carriers of the mutant allele collectively resulted in the determination that individuals harbouring the mutant allele presented with a significantly higher mean two-hour plasma insulin concentration (p value < 0.04). In a subset, separate from these individuals, it was determined the mutant allele is significantly associated with lower levels of insulin

stimulated glucose uptake as well as increased fatty acid oxidation. It was further determined by Baier *et al.* that the mutation was associated with an increased affinity of FABP2 towards fatty acids therefore ensuring the preferential uptake of these molecules. Chronically increased levels of fatty acids in turn result in increased triglyceride accumulation within various non-adipocytes such as β cells and myocytes (Unger *et al.*, 1999). Although this mutation has been associated with numerous clinical factors involved in insulin resistance, it was however not associated to T2D within the Pima Native American cohort investigated.

The Ala54Thr mutation was also investigated within a Native Canadian population (Hegele *et al.*, 1996) where harbouring the mutant allele was significantly associated with increased BMI (p value = 0.012), increased % body fat (p value = 0.019) and increased plasma triglycerides (p value = 0.012). The allele was however neither associated with T2D disease risk or hyperinsulinaemia as it was in the Pima Native Americans. It was determined that the Canadian cohort investigated was on average five years older than the Pima cohort. It is thus possible that the increased insulin secretion induced by the Ala54Thr mutation in the Pima Native Americans could no longer be maintained in the older native Canadian cohort, due to β cell loss caused by the over-secretion (Hegele *et al.*, 1996).

Further association of the mutant allele with increased plasma triglycerides was reported in diabetic Caucasians (Georgopoulos *et al.,* 2000). Patients harbouring the mutant allele presented with significantly higher fasting levels of plasma triglycerides for heterozygotes (p value = 0.0007) and for the mutant allele homozygotes (p value < 0.000007). Upon comparison of a subset of these patients, one group harbouring the wild type allele and another harbouring the mutant allele, it was determined that following a high fat meal the patients harbouring the mutant allele had significantly increased (p value = 0.025) levels of plasma triglycerides (Georgopoulos *et al.*, 2000)

It therefore may be hypothesised that although not directly responsible for inducing T2D, variation at this locus does however mitigate the pathogenesis of the disease. The increased levels of plasma triglycerides (Georgopoulos *et al.*, 2000) and the concomitant obesity (Hegele *et al.*, 1996) are both known risk factors for CAD, which is the lethal counterpart to T2D (Reavan, 1988). Thus the FABP2 locus, although not necessarily important in causing T2D, is indeed relevant to disease progression.

3.1.11 The tumour necrosis factor alpha (TNFα) gene

Analysis of a native Canadian population with regard to possible relationships between TNFα and various physiological and anthropometric variables resulted in the determination of an association between circulating levels of this cytokine and insulin resistance. It was furthermore determined that high levels of TNFα were linked to increased systolic blood pressure (Zinman *et al.*, 1999).

Mutations affecting the expression level of TNFα are associated with T2D via the aforementioned mechanism (Hotamisligil and Spiegelman, 1994) as presented in Section 2.1.2.2. An example of this is the alteration termed G-308A, which results in increased expression of the cytokine consequently causing insulin resistance and greater BMI in patients harbouring the alteration. This has been hypothesised to be due to the elevated leptin levels caused by the aforementioned mutated protein (Fernández-Real *et al.*, 1997).

3.1.12 The interleukin-6 (IL6) gene

The association of TNFα to insulin resistance has led to the investigation of the association of other cytokines to this pathogenic state. It was elucidated that interleukin-6 (IL6) levels were increased in the plasma of individuals with T2D (Bastard *et al.*, 2002). It was furthermore determined that this increase was not due to increased cytokine production via the immune system (Pickup *et al.*, 2000), but as the levels of IL6 were significantly associated with BMI (p value < 0.01) it was hypothesised that the major source of this increase was due to increased expression in adipocytes (Bastard *et al.*, 2002). Further analysis allowed for the determination that the IL6 content of adipocytes are inversely proportional to insulin stimulated glucose uptake both *in vivo* as well as within the adipocytes themselves.

Analysis of an insulin-resistant murine model allowed for the discovery that upon insulin stimulation, a marked increase in IL6 expression could be distinguished in skeletal muscle in comparison to control individuals (Carey *et al.*, 2003). It is therefore possible that the increased plasma IL6 levels originate in both these tissues in insulin resistant individuals.

In an investigation by Stephens *et al.* (2004) of a promoter SNP within the IL6 gene termed G-174C, it was discovered that diabetics harbouring the variant allele in either the heterozygotic or homozygotic form were significantly more obese (p value $= 0.05$) as compared to non-carriers. This alteration has also previously been associated with increased blood pressure and CAD in individuals harbouring the heterozygous form

(Humphries *et al.*, 2001). The association to obesity was however not maintained upon investigation of a group of non-diabetic controls. Upon comparison of the two groups it was determined that there were more homozygotes for the variant allele present within the non-diabetic group. The apparent lack of homozygotes in the diabetic cohort lead Stephens *et al.* to hypothesise that diabetic individuals harbouring this genotype had a higher level of fatality, as the CAD risk imparted due to insulin resistance (Reavan, 1988) was increased in these individuals due to the effects of the alteration as discussed by Humphries *et al.* (2001). Future analyses of the mechanism by which this alteration increases disease risk are required to elucidate the underlying pathway in CAD and T2D risk.

3.1.13 The hepatocyte nuclear factor 4 alpha (HNF4α) gene

HNF4α was localised to chromosome 20q12-q13.1 via linkage analysis and *in situ* hybridisation (Argyrokastritis *et al.*, 1997). It was originally described with reference to MODY susceptibility however it is becoming apparent that this gene harbours alterations which may be involved in increased risk towards T2D. An investigation into a group of 19 French Caucasian late-onset T2D families resulted in the determination of three intronic alterations and a single alteration within the coding region of a specific family termed Val393 isoleucine (Ile). It was found to segregate with the disease and was nominally associated (p value = 0.043) with a decrease in insulin secretory function in family members that harboured the alteration but not the disease. Upon analysis of the transactivation ability of this mutated TF it was discovered that this activity was decreased by at least 50%. However due to the fact that this alteration was determined in a single family it is unlikely that it is a significant risk factor in T2D unless it is subsequently also found in other families (Hani *et al.*, 1998). As will be discussed, intronic sequences may be important in T2D disease susceptibility (Horikawa *et al.*, 2000) and therefore it may be possible that the intronic variants determined in the investigation by Hani *et al.* may harbour genetic risk.

3.1.14 The hepatocyte nuclear factor 1 alpha (HNF1α) gene

The human HNF1α gene was localised to chromosome 12 through the utilisation of hamster-human somatic cell hybrids (Kuo *et al.*, 1990). The map position of this locus was refined to chromosome 12q22-terminal region (ter) via hybrid mapping (Kuo *et al.*, 1990).

Numerous mutations within this gene have been associated with MODY3 (Yamagata *et al.*, 1996b; Vaxillaire *et al.*, 1997), however it has also been associated with T2D (Hegele *et al.*, 1999). Upon screening of the HNF1α gene within the Oji-Cree aboriginal Canadians (Hegele *et al.*, 1999) the Gly319Ser alteration was detected and presented with significant association to T2D (p value = 0.000001). Upon odds ratio (OR) analysis it was determined that homozygotes of the mutant allele had a value of 4 (95% confidence interval (CI) 2.64-6.03) indicating an increased risk for T2D when harbouring this allele. Further analysis however indicated a significant association between earlier age of onset and the presence of the variant allele (p value = 0.009 for heterozygotes and p value = 0.001 for homozygotes of the variant allele). BMI was also lower in individuals harbouring the mutant allele, in either the heterozygous (p value = 0.058) or homozygous forms (p value = 0.017). Thus, although not MODY, this alteration may cause a different form of early onset T2D (Hegele *et al.*, 1999). As this alteration was only found in the Oji-Cree cohorts investigated it may represent a population specific risk factor that requires further elucidation of its mechanism of action.

Analysis of transcriptional activity of the HNF1α protein harbouring the Gly319Ser mutation allowed for the elucidation of the possible mechanism of action (Triggs-Raine *et al.*, 2002). It was determined that the presence of the mutant allele resulted in a 54% decrease in transcriptional activity (p value < 0.0001). This decrease was however not due to impaired DNA binding or dimer stability and therefore it was postulated that this effect was due to decreased transcription following DNA binding. It was further hypothesised that individuals harbouring this allele had compromised insulin secretion, which was possibly aggravated by the effects of obesity or other T2D susceptibility loci thus, placing the patients at an increased risk of developing the early onset form of T2D associated with this mutation in the Oji-Cree (Triggs-Raine *et al.*, 2002).

3.1.15 The hepatocyte nuclear factor 1 beta (HNF1β) gene

The gene encoding HNF1β was localised to chromosome 17q via the screening of a human-rodent somatic cell hybridisation panel utilising PCR (Abbott *et al.*, 1990). This locus was further refined via *in situ* hybridisation to chromosome 17q11.2-q21.1 (Bach *et al.*, 1991).

An investigation of a Japanese patient population resulted in the determination of two alterations possibly involved in diabetic susceptibility (Furuta *et al.*, 2002). The first was

only observed in a single patient but upon functional analysis it was determined that this mutation (Arg276X) resulted in the complete ablation of HNF1β transactivation activity. This alteration was, however, detected in an individual who had been misdiagnosed with T2D but was actually suffering from MODY5, as was determined by low insulin secretory ability, diminutive kidneys and an age of onset of 13 years. The second alteration (Ser465Arg) was detected in two late onset T2D patients both of whom did not harbour any form of renal disease. Upon testing the mutated protein's transactivation ability, it was determined to be 88% of that of the wild type form (Furuta *et al.*, 2002). Thus it is possible that alterations within the HNF1β gene can result in MODY or T2D depending on the severity of the mutational effect.

3.1.16 The insulin promoter factor 1 (IPF-1) gene

IPF-1 was localised to chromosome 13q via the use of human/rodent somatic cell hybrids and was further defined to be present on 13q12.1 via the application of FISH analysis (Stoffel *et al.*, 1995). As this gene was determined to be involved in susceptibility to MODY (Ohlsson *et al.*, 1993) it became an important candidate for T2D susceptibility.

It was determined by Macfarlane *et al.* in 1999 that depending on the severity of the mutation, it is possible for an individual to present with MODY or T2D. This was achieved by first screening the IPF-1 gene in individuals suffering from MODY, who had the various other MODY genes previously excluded, in order to determine possible alterations. In turn diabetic patients were screened to determine if the various alterations were present. The authors discovered three missense mutations namely a cysteine (Cys) 18 Arg, an Asp76Asn and an Arg197 histidine (His). All three mutations were present in affected individuals although this was not significant due to the relative rarity of the risk alleles. A relative risk of 3.0 (range: 1.0 - 9.2; p value = 0.04) was determined by Macfarlane *et al.* upon odds ratio analysis of the T2D probands. To determine if the mutations resulted in a decrease in functionality, each was compared to normal IPF-1 with regard to three specific abilities. The first was the ability to translocate to the nucleus within the β cell, in a phosphorylation dependent manner upon glucose stimulus. It was determined that this was unaffected by the various mutations. Secondly, the binding activity to the insulin promoter was investigated and it was determined that the Cys18Arg, Asp76Asn and Arg197His mutations resulted in a decrease in this activity with the greatest effect being in the case of the Arg197His mutation. The final functional test was undertaken in order to determine if there was any variation of insulin gene transcription in response to the various

mutations. It was determined that similarly to the binding ability, this activity was also decreased with the largest decrease in the case of the Arg197His mutation. The increased effect of the Arg197His mutation is possibly due to the fact that it is present within the DNA-binding domain of IPF-1 (Macfarlane *et al.*, 1999). It is evident from the various functional and association studies that these missense mutations definitely increase susceptibility to T2D. However, their rarity implies that they are not major contributors to the genetic aetiology of this disorder.

Further analyses allowed for the recognition of two more mutations, which could have been responsible for causing T2D, within the French Caucasian population namely a substitution of a glutamine (Gln) 59 leucine (Leu) as well as an insertion of a proline encoding codon at nucleotide position 243 (insCCG243). The Asp76Asn was also detected in this investigation (Hani *et al.*, 1999). Upon association analyses via a case-control investigation it was elucidated that Asp76Asn was associated with T2D (relative risk = 12.9). Upon comparison of the OGTT of individuals harbouring the Asp76Asn mutation to that of normoglycaemic individuals, it was discovered that the mutation carriers had significantly lower insulin levels in response to the glucose challenge. The Gln57Leu mutation was deemed to have an additive effect on T2D risk although it was rather rare within the population investigated. The insCCG243 mutation however was linked to a seemingly autosomal dominant form of late onset T2D, and induced a progressive decline in insulin secretion. It is therefore unlikely that it is involved in the polygenic aetiology of T2D (Hani *et al.*, 1999). IPF-1 may be an important additive locus in T2D susceptibility interacting with the cumulative effect of insulin resistance.

3.1.17 The glucokinase gene

As stated in Section 2.1.4.3.1.1 the glucokinase gene was localised to chromosome 7p (Froguel *et al.*, 1992). It has been determined to consist of twelve exons as elucidated via sequence determination of four clones, identified through screening of a human placental genomic library, utilising labelled cDNA of the human glucokinase gene (Stoffel *et al.*, 1992).

Although described with reference to MODY2 this gene has also been hypothesised to be involved in susceptibility to T2D (Froguel *et al.*, 1992). This is due to the central role that the protein product has in glucose sensing (Matschinsky, 1990). Investigation of this gene as a possible candidate has resulted in the elucidation of various alterations associated

with T2D risk. Investigation of the genotype distribution at the G-258A locus in a normoglycaemic Asian Indian cohort allowed for the determination that harbouring the A allele decreased the sensitivity of the liver towards insulin (Chiu *et al.*, 2000). The function of the β cells in individuals with the A allele was however upregulated, therefore masking this decreased hepatic sensitivity (Chiu *et al.*, 2000). This is an excellent example of the pathway by which an underlying deficiency can result in greater susceptibility to T2D, as the hyperinsulinaemia generated by the β cells will eventually result in the faster degradation of these cells.

A second promoter polymorphism termed G-30A was examined via a case-control study for CAD (Nauck *et al.*, 2003). The A allele was associated with increased risk towards CAD (OR 1.34, 95% CI 1.12-1.60) as well as T2D (OR 1.20, 95%CI 1.03-1.40). Individuals harbouring the mutant allele also presented with significantly increased plasma glucose levels during an OGTT with a p-value of 0.002 at fasting and a p-value of 0.017 at 120 minutes. This was most likely due to the decreased β cell function associated with this allele (p-value = 0.012). Thus it is important to not only screen for exonic variants in investigating T2D susceptibility as intronic alterations, as well as extragenic factors, can also modulate disease risk.

3.1.18 The ras associated with diabetes (rad) gene

The ras associated with diabetes (*rad*) gene was localised to chromosome 16q22 through the utilisation of FISH analysis (Doria *et al.*, 1995). The utilisation of restriction fragment melting polymorphism (RFMP) via denaturing gradient gel electrophoresis (DGGE) allowed for the elucidation of a RFMP encompassing the *rad* locus termed RAD1, a RFMP with 10 allelic forms generated from variable GTT and ATT repeating sequences (Doria *et al.*, 1995). The various allele compositions are presented in Table 3.1.

A control cohort (n=133) and a diabetic cohort (n=210) were screened at this locus wherein one specific allele (allele 8) was present at >80% within both groups (Doria *et al.*, 1995). It was however determined that certain minor alleles were present at a greater frequency within the T2D patients. Upon pooling of alleles 1, 2, 5, 6 and 9 and comparison of this group with the remaining five allele types utilising χ^2 analyses, it was determined that the minor alleles within the pooled group were significantly associated with disease risk (p value < 0.025). The alleles were pooled into four classes as indicated in the final column of Table 3.1 according to the GTT repeats followed by the number of ATT repeats.

Utilising this classification it was determined that Class I, II and IV were significantly more prevalent in T2D individuals (p value < 0.001).

Table 3.1: Composition of alleles present at the RAD1 locus within the *rad* gene

Allele number	$(GTT)_n$	$(ATT)_n$	Number of repeats	Class	Allele number	$(GTT)_n$	$(ATT)_n$	Number of repeats	Class
1	5	14	19	I	**6**	6	8	14	II
2	5	13	18	I	**7**	5	8	13	III
3	5	12	17	I	**8**	5	7	12	III
4	6	10	16	II	**9**	5	6	11	III
5	6	9	15	II	**10**	4	5	9	IV

Adapted from Doria *et al.* (1995).

This investigation by Doria *et al.* was undertaken in Caucasian Americans, however, it was necessary to determine if a similar pattern was present within other Caucasian populations. Orho *et al.* (1996) thus screened a Finnish diabetic (n=290) and a control cohort (n=270) for the allele distribution at this locus. Similar to the investigation by Doria *et al.* (1995) allele 8 was present in ca. 80 % of the cohorts investigated, however no significant association could be detected for any of the other alleles. Upon comparison of the classes as previously defined no significant association was determined within the cohorts investigated (Orho *et al.*, 1996). This locus is thus an excellent example of possible population specific effects of susceptibility loci.

3.2 MURINE MODELS OF TYPE 2 DIABETES MELLITUS

Another useful strategy in the elucidation of risk loci in T2D is the use of animal models, specifically rodents. Various murine knock out strategies have been used to generate models of specific monogenic forms of T2D (Kido *et al.*, 2000). In addition, naturally occurring forms of the disease within rodents also indicate towards possible loci of importance (Maffei *et al.*, 1995).

3.2.1 Obese/obese

Upon analysis of the genomic organisation of the *obese* gene in mice it was determined that, similar to the human gene, it consists of three exons and two introns (He *et al.*, 1995). The murine gene furthermore harbours an alternate exon within the first intron, which through alternative splicing is present within ca. 5% of mRNA transcripts within adipose

tissue. The promoter region harbours a TATA binding element and is able to bind C/EBP as was predicted in the human sequence (He *et al.*, 1995).

3.2.2 Diabetic/diabetic rat

This diabetic/diabetic (db/db) murine model presented with increased levels of the *obese* gene mRNA transcripts (Maffei *et al.*, 1995). Furthermore it was determined that hypothalamic lesions induced a similar phenotype thus indicating that both the db gene product, as well as the hypothalamus, play important roles in the murine signalling pathway of leptin (Maffei *et al.*, 1995).

3.3 TYPE 2 DIABETES MELLITUS SUSCEPTIBILITY LOCI DETERMINED VIA LINKAGE

The application of the more classic monogenic tools such as linkage testing to T2D susceptibility has allowed for the elucidation of various loci within the genome, which may be responsible for disease risk (Hanis *et al.*, 1996). This has resulted in the discovery of CAPN10 (Horikawa *et al.*, 2000) and APM1 (Kissebah, 2000) neither of which were originally candidates for disease risk, thus highlighting the relevance of this strategy.

3.3.1 The calpain 10 (CAPN10) gene

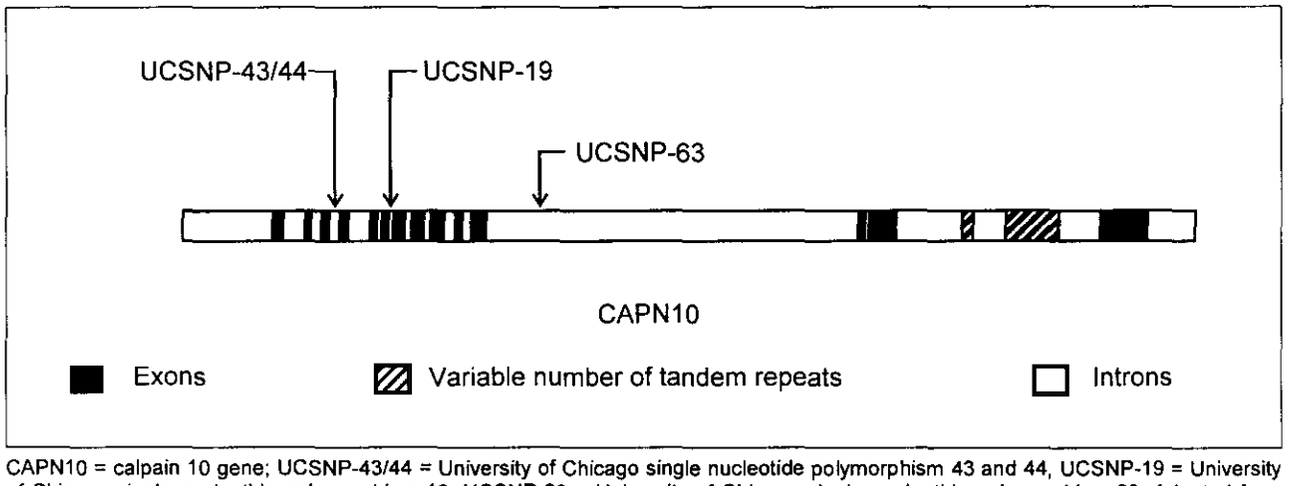
Following a genome wide scan for T2D susceptibility loci within the Mexican American population a LOD_{max} score of 2.58 (p value < 0.0005) was detected for a marker localised to chromosome 2q. In turn the locus was designated non-insulin dependent diabetes mellitus locus 1 or NIDDM1 (Hanis *et al.*, 1996). The gene expressing the protein CAPN10 is located at this specific locus and is a ubiquitously expressed protease. Various polymorphisms present in this gene have been associated with T2D (Horikawa *et al.*, 2000).

Calpains are calcium dependent intracellular non-lysosomal proteases, which cleave proteins at a minimum number of sites. The fact that many cellular functions are controlled by calcium as well as the fact that these proteins do not degrade, but rather modify other proteins, leads to the hypothesis that these proteases are regulatory in nature (Croall and Demartino, 1991). Although the calpain family of proteins have been described in the pathogenesis of a number of disorders, it is CAPN10, which has been determined to be the most important with regard to T2D. This caspase-like cysteine protease is an example of reverse genetics whereby the locus was determined to be associated with T2D, but to date its role in disease pathogenesis has not been elucidated (Horikawa *et al.*, 2000).

Although the exact role of this protein in disease pathogenesis is currently unknown, a possible mechanism of action is discussed in Section 6.2.

The reported alterations within the CAPN10 gene are all SNPs found in intronic regions. The polymorphisms have been designated as University of Chicago single nucleotide polymorphisms (UCSNP) with reference to the institution where these alterations were first characterised. UCSNP-43 and UCSNP-44 are both located in intron 3, UCSNP-19 is present in intron 6 and UCSNP-63 in intron 13 of the aforementioned gene. The positions of these alterations within the CAPN10 gene are depicted in Figure 3.1. It is hypothesised that these polymorphisms affect the binding of various transcription factors, thereby altering the expression profile of this gene, ultimately leading to the diseased phenotype (Horikawa *et al.*, 2000).

Figure 3.1: Diagrammatic representation of the calpain 10 gene structure



CAPN10 = calpain 10 gene; UCSNP-43/44 = University of Chicago single nucleotide polymorphism 43 and 44, UCSNP-19 = University of Chicago single nucleotide polymorphism 19; UCSNP-63 = University of Chicago single nucleotide polymorphism 63; Adapted from Horikawa *et al.* (2000).

3.3.1.1 University of Chicago single nucleotide polymorphism (UCSNP)-43

UCSNP-43 is located in intron 3 of the CAPN10 gene and is a guanine to an adenine alteration at nucleotide (nt) position 4852 of which the G allele has been associated with an increased risk to T2D in the Mexican American population (Horikawa *et al.*, 2000). It has also been determined in an African American cohort that UCSNP-43 is associated with T2D susceptibility (Garant *et al.*, 2002). In this population the "at-risk" G/G genotype is present at a much higher frequency (allele frequency = 0.9) than in the Caucasian populations (see Table 3.2 for comparison of non-African populations). This SNP, however, has not been associated with any variation in the secondary symptoms of T2D such as changes in BMI, waist to hip ratio, fasting serum and insulin levels. It has been

hypothesised that this is due to UCSNP-43 interacting with a locus on chromosome 15 therefore indirectly exacting its effect on the affected phenotype of the patient (Garant *et al.*, 2002). Definitive association of this allele with disease susceptibility in various Caucasoid populations has not yet been determined.

Table 3.2: Relative frequencies of the G allele of UCSNP-43 in different non-African populations

Population	South Indian	United Kingdom	German	Polish	Mexican Americans	Pima Native Americans
Allele frequency	0.81[a]	0.73[b]	0.70[c]	0.73[d]	0.80[c]	0.62[e]

a = Cassell *et al.* (2002); b = Evans *et al.* (2001); c = Horikawa *et al.* (2000); d = Malecki *et al.* (2002); e = Baier *et al.* (2000).

Within the Native American population of the Pima Native Americans, Baier *et al.* (2000) determined that the 'at risk' G/G genotype resulted in increased susceptibility, as individuals harbouring this genotype presented with an increased plasma glucose level. It was concluded that this increase was due to a lower level of glucose turnover caused by lower rates of glucose oxidation. An interesting corollary to this is that lipid oxidation was increased in individuals harbouring this alteration. Investigation of CAPN10 mRNA levels within the skeletal muscle of this population allowed for the determination that individuals with the G/G genotype presented with lower mRNA levels. This was one of the first indications for the mechanism by which this alteration causes increased T2D susceptibility (Baier *et al.*, 2000).

3.3.1.2 University of Chicago single nucleotide polymorphism (UCSNP)-44

UCSNP-44 is a thymine to a cytosine alteration at nt position 4,841. The reason for the increased susceptibility of patients harbouring this SNP is not yet known. The relative frequencies of this alteration within the various populations in which it has been defined are presented in Table 3.3. UCSNP-44 is generally not included in the haplotype structure determined in many populations as its differentiation power within the haplotype is not as strong as the other SNPs discussed (Horikawa *et al.*, 2000).

Table 3.3: Relative frequencies of the T allele of UCSNP-44 in different non-African populations

Population	South Indian	United Kingdom	Mexican Americans
Allele frequency	0.85[a]	0.84[b]	0.90[c]

a = Cassell *et al.* (2002); b = Evans *et al.* (2001); c = Horikawa *et al.* (2000).

3.3.1.3 University of Chicago single nucleotide polymorphism (UCSNP)-19

UCSNP-19 is located within intron 6 and is an insertion deletion mutation of 32 bp at nt 7,920. The "at-risk" allele, containing three 32 bp repeats, has been associated with a decrease in insulin sensitivity within the Caucasoid population (Elbein *et al.*, 2002). This was determined due to the fact that the fasting level of insulin is increased in patients harbouring the "at-risk" allele. The various allele frequencies of this alteration within the diabetics of specific populations are presented in Table 3.4.

Table 3.4: Relative frequencies of the two repeat allele of UCSNP-19 in different non-African populations

Population	South Indian	United Kingdom	German	Polish	Mexican Americans
Allele frequency	0.41[a]	0.38[b]	0.66[c]	0.34[d]	0.58[c]

a = Cassell *et al.* (2002); b = Evans *et al.* (2001); c = Horikawa *et al.* (2000); d = Malecki *et al.* (2002).

3.3.1.4 University of Chicago single nucleotide polymorphism (UCSNP)-63

Intron 13 harbours UCSNP-63, which is a transition alteration known as C16378T. The "at risk" allele of this polymorphism is very rare within the Caucasoid population, however, it has been concluded that in association with UCSNP-19, this alteration is associated with an increase in insulin resistance thus increasing T2D susceptibility (Elbein *et al.*, 2002). The frequency at which the unaffected allele is present in the various populations is depicted in Table 3.5.

Table 3.5: Relative frequencies of the C allele of UCSNP-63 in different non-African populations

Population	South Indian	United Kingdom	German	Polish	Mexican Americans
Allele frequency	0.97[a]	0.92[b]	0.94[c]	0.93[d]	0.77[c]

a = Cassell *et al.* (2002); b = Evans *et al.* (2001); c = Horikawa *et al.* (2000); d = Malecki *et al.* (2002).

3.3.1.5 University of Chicago single nucleotide polymorphism (UCSNP) haplotypes

Specific haplotypes of these four SNPs have been associated with different symptoms, which increase T2D susceptibility in various populations. The order in which the four SNPs are presented in the haplotypes, is as follows: UCSNP-44, UCSNP-43, UCSNP-19 and UCSNP-63. Within the South Indian population the presence of the 1112/1121 heterozygous haplotype was associated with an increase in the risk of impaired fasting

glucose/impaired glucose tolerance with a factor of ca. 10 (Cassell *et al.*, 2002). This haplotype is also hypothesised to be responsible for a 6.3-fold increase in T2D risk in patients that are not related and a 5.8-fold increase in the probands of the investigation undertaken by Cassell *et al.*

The 1112/1121 haplotype is equivalent to the 112/121 haplotype in the Mexican American population as presented by Horikawa *et al.* (2000), with the inclusion of UCSNP-44 in the South Indian population being the only difference. It has been established that this haplotype combination is responsible for a 3-fold increase in T2D risk in the aforementioned Mexican American population.

In an investigation of patients affected by polycystic ovary syndrome (PCOS), a disorder presenting with similar symptoms to T2D, it was determined, that in the African American cohort the 112/121 haplotype was associated with a significant increase in insulin levels in response to the OGTT test (Ehrmann *et al.*, 2002). The haplotype combination was also related to a 2-fold increase in PCOS risk.

It has however been determined that there is large variation in the prevalence of at risk alleles and specific haplotype combinations, depending on which population is under investigation. The homozygous 121 haplotype has been discovered to be more prevalent in diabetics from the Polish population. It was further determined that the 112/121 haplotype was not responsible for increased risk in the Polish population (Malecki *et al.* 2002).

Furthermore, slight association has been determined between these aforementioned haplotypes and disease susceptibility within the Samoan population. Diabetic individuals presented with a slight increase in frequency of the 112/121 and 121/221 haplotypes these were however, not significant (Tsai *et al.*, 2001).

These investigations lend credence to the hypothesis that CAPN10 only influences risk to diabetes susceptibility in certain populations (Tsai *et al.*, 2001). This would be expected, as natural selection cannot act as strongly on susceptibility loci as it does on loci that have direct positive or negative effects, thus resulting in more variation within these various aforementioned loci in the evolution of humankind. It is still necessary to determine in which populations CAPN10 has an effect in order to develop an actual model of T2D susceptibility in these specific populations.

With the aforementioned in mind it has been determined that there are large genotype and haplotype frequency variations between the African and non-African populations. In the Mbuti and Biaka populations of Central Africa, the 112 haplotype (frequency = 0.59 in Biaka and 0.82 in Mbuti) is much more common than the 121 haplotype (frequency = 0.00 in both cohorts) which has been found to be most prevalent in non-Africans. The African populations investigated by Fullerton *et al.* (2002), did not contain the 112/121 "at risk" haplotype and the authors hypothesised that CAPN10 may not be as important in diabetes susceptibility in the African population (Fullerton *et al.*, 2002). It can however be argued that due to different selection criteria in terms of diet, activity and metabolism for these populations, alternative haplotype combinations as compared to the non-African populations may impose greater risk.

3.3.2 The non-insulin dependent diabetes mellitus 2 (NIDDM 2) locus

The non-insulin dependent diabetes mellitus 2 (NIDDM2) locus was identified in a genome wide scan within Finnish T2D families, which had been enriched for low levels of insulin secretion (Mahtani *et al.*, 1996). The initial investigation included all T2D families, irrespective of their insulin secretory status however no linkage was determined. It was only upon segregation of families into groups based on insulin secretion capability that borderline significant association to chromosome 12q was determined (Z_{all} = 4.1). As this is also the site of the MODY3 locus it was hypothesised that this locus may result in MODY or T2D dependent on the specific alleles that were inherited (Mahtani *et al.*, 1996).

The search for T2D susceptibility loci has lead to the utilisation of various genes involved in the monogenic forms of the disease i.e. MODY as candidates for linkage analyses. Shaw *et al.* in 1998 utilised this reasoning to delineate a region for investigation in a pedigree presenting with late onset T2D with severe insulin resistance and an autosomal dominant mode of inheritance. The investigators utilised four markers within the MODY3 region and via multipoint linkage analysis Z_{all} of 3.65 was generated at a locus on chromosome 12q. The 10 exons and the promoter region of the HNF1α gene was subsequently screened for any alterations that could have been responsible for the Z_{all} score. No association between any of the alterations at this locus and the T2D phenotype were detected (Shaw *et al.*, 1998). Linkage to this locus was however only detected in a single pedigree and therefore its effect may not be relevant in the greater population. Although the pedigree utilised was advantageous in that it presented with an autosomal

dominant inheritance pattern, this in itself would bias the experiment towards individuals with severe insulin resistance, which, as discussed previously, may have a different pathogenesis that is distinct from T2D.

An analysis of chromosome 12q in 26 Caucasian and 6 non-Caucasian families presenting with early onset autosomal dominant T2D by Bektas *et al.* in 1999 resulted in the determination of Z_{all} 50 centiMorgans (cM) closer to the centromere than NIDDM2 (Z_{all} = 2.9, p value = 0.015). Upon removal of the non-Caucasian families from the analysis there was an increase in the Z_{all} value i.e. 3.8 (p value = 0.007) which would be explained by the removal of the inter population genetic heterogeneity. It was further stated that the absence of linkage to NIDDM2 might have been due to the utilisation of pedigrees with early onset T2D that may have biased the analysis, especially if the NIDDM2 locus is more closely linked to the late onset form of the disease (Bektas *et al.*, 1999). The role of this locus in T2D susceptibility is therefore marred as the utilisation of pedigrees presenting with the autosomal dominant forms of the disease, is an artificial method for determining T2D susceptibility loci as these two forms of the disorder can be seen as having different aetiologies. It therefore becomes necessary to describe the T2D phenotype definitively, which is often difficult with the current available tools. However, it is an integral step in the process of determining the genetic aetiology of this group of disorders.

3.3.3 The non-insulin dependent diabetes mellitus 3 (NIDDM 3) locus

The non-insulin dependent diabetes mellitus 3 (NIDDM3) locus was determined following the discovery that certain candidate genes mapped to chromosome 20q. These included genes which encoded phosphoenolpyruvate carboxykinase (PEPCK), overexpression of which has been linked to T2D development in mice (Valera *et al.*, 1994), HNF4α which has been linked to MODY1 (Yamagata *et al.*, 1996a) and agouti signalling protein (ASP) which is the human orthologue to a gene involved in the obesity-diabetes phenotype in a specific murine model (Bell *et al.*, 1991). Due to this high proportion of candidate genes on chromosome 20q, Zouali *et al.* in 1997 attempted to determine if the diabetic phenotype displayed linkage to this chromosomal region.

In this investigation 301 sib pairs from 148 families of the French Caucasian population were utilised in a linkage study. Within this group, a subset of individuals (55 sib pairs from 42 families) with an age of diagnosis of less than 45 years, were separately investigated for linkage disequilibrium. It was hypothesised that individuals, who presented with T2D at

a younger age, would possibly harbour a stronger genetic predisposition. Upon multipoint analysis of 10 polymorphic markers, with an average density of 7.5 cM, it was determined that T2D presented with strong linkage (LOD_{max} = 2.74, p value = 0.0004) to a position 7 cM proximal to the PEPCK gene (confidence interval = 24 cM) within the early onset T2D subset. Upon utilisation of a conservative weighting procedure this LOD score remained significant (LOD_{max} = 2.34, p value = 0.0009).

Analysis of the entire set of affected families resulted in the detection of linkage at the ribophorin II (RPNII) gene (LOD_{max} = 1.81, p value = 0.0003) with a confidence interval of 28 cM including the region termed 20q12-13.11. The locus thus includes the genes encoding phospholipase C (PLC) as well as adenosine deaminase (ADA). Zouali *et al.* (1997) therefore hypothesised that as the subset of early onset T2D linked to the proximal region of the PEPCK gene, any alterations responsible for the increased risk would be within the upstream region of this gene. Upon SSCP analysis it was determined that there were no polymorphisms within the individuals screened. However this form of analysis is not 100% sensitive (Hayashi and Yandell, 1993) and thus it is possible that an alteration may be present but was just not detected via this methodology.

The apparent lack of causative alterations within the PEPCK gene was one of the factors, which lead Price *et al.* (1999) to determine a physical map for the region spanning 20q12-q13.1 in order to define any possible candidates responsible for the linkage to this domain. The resulting physical map included such candidates as phospholipid transfer protein, syndecan 4, receptor for activated C-kinase type 7 (RACK7) and the CCAAT enhancer binding protein B (C/EBP B). A YAC / bacterial artificial chromosome (BAC) contiguous domain of the region which could be utilised in future investigations of T2D susceptibility loci was also produced (Price *et al.*, 1999). RACK7 was one of the initial candidate genes determined via this method. The gene was rapidly mapped utilising the physical map, while its genomic organisation was unravelled utilising the YAC/BAC contiguous clone set. Following mutation analysis (SSCP) however it was determined that this gene did not play a significant role in susceptibility towards T2D (Fossey *et al.*, 1998).

In the Finnish population, further evidence was determined indicating that chromosome 20 harboured a T2D susceptibility locus. Ghosh *et al.* in 1999 undertook the screening of 716 affected sib pairs from 477 families for linkage to T2D. Three LOD_{max} scores were determined at 18.5 cM (LOD_{max} = 1.92, p value = 0.013), 57 cM (LOD_{max} = 2.06, p value = 0.009) and 69.5 cM (LOD_{max} = 2.00, p value = 0.01) according to the map

location presented by Ghosh *et al.* (1999). When the cohort was clinically stratified it was determined that certain families were responsible for a disproportionate amount of the detected linkage. Upon further clinical subdivision it was determined that families with the lowest BMI (3% of the sibships investigated) had a non-significant LOD score of 2.56 at 53 cM. Patients with the highest ratio of fasting c-peptide to fasting glucose (4% of the sibships investigated) produced a LOD_{max} of 3.46 at 53.5 cM however this was not statistically significant. The authors thus posited that these subgroups might explain the linkage peak at 57 cM. Low BMI however is generally associated with a decreased risk towards T2D and thus it seems unlikely to be linked to a locus of T2D susceptibility. If it is taken into account that these calculations were also non-significant it becomes evident that this hypothesis still requires further testing. Similarly by further subdividing the cohort, certain clinical factors allowed for the elucidation of LOD_{max} scores close to those regions detected via full cohort investigation i.e. sibships with the highest ratio of fasting insulin to fasting glucose (17% of the sibships investigated) produced a LOD_{max} score of 3.06 at 66 cM while those with highest mean fasting c-peptide levels (19% of the sibships investigated) yielded a LOD_{max} score of 2.93 at 21 cM.

Due to the previous association of the MODY1 gene HNF4α (Yamagata *et al.*, 1996a) it was hypothesised by Ghosh *et al.* (1999) that the LOD_{max} determined at 69.5 cM was due to alterations within the aforementioned gene, which had been mapped to a position of 62.7-66.1 cM. Upon molecular screening of the twelve exons and the promoter region of 16 controls and 64 affected individuals, 14 base alterations and three deletions were detected. Five of these base alterations were rare i.e. specific to certain diabetic families, however each specific alteration was present within affected members of the families and absent in unaffected family members. Of the six newly described common changes not one associated with increased T2D risk. The remaining four alterations had been previously reported but also did not present with any association. In fact the thymine to cytosine change at nucleotide position 38 in intron 1b was significantly more prevalent in the control population investigated. Upon removal of the families harbouring the rare variants, the LOD_{max} at 69.5 cM dropped to 1.74 thus implying that the effect of these families biased linkage analysis thus inflating a LOD_{max} score and therefore increasing its significance. Ultimately Ghosh *et al.* (1999) presented evidence for the presence of two possible diabetes susceptibility loci on chromosome 20.

3.3.4 The adiponectin (APM1) gene

In the investigation of quantitative trait loci (QTL) involved in the susceptibility towards the metabolic syndrome it was determined that a locus on 3q27 was strongly linked to six traits fundamental to the expression of this cluster of disorders. These traits include BMI, waist circumference, fasting plasma insulin, body weight, hip circumference and the ratio of glucose to insulin. Secondary to this association it was determined that this locus displayed epistasis with another QTL on chromosome 17p12, which was in turn closely associated with plasma leptin levels (Kissebah, 2000).

The gene encoding APM1 has been localised to chromosome 3q27 and is ca. 16 kb in length. The genetic structure consists of two introns (12 kb and 800 bp respectively) as well as three exons (18 bp, 222 bp and 4,277 bp respectively). The introns contain typical GT-AG splice junctions (Saito *et al.*, 1999). The 5'-untranscribed region contains numerous alternate promoter elements as opposed to the general TATA promoter element. The specificity of expression by adipocytes may be due to the C/EBP elements present within this region.

There are both environmentally and genetically induced means by which hypoadiponectinaemia can originate (Fasshauer *et al.*, 2001; Kondo *et al.*, 2002). An important example by which the environment may affect APM1 levels is through the action of catecholamines. It has been determined that catecholamines are important in controlling the expression levels of APM1. Catecholamines induce down regulation of APM1 via a β-adrenergic receptor, which in turn activates a protein kinase A dependent pathway through the action of a stimulatory guanine nucleotide-binding protein. It has been hypothesised that this is the process by which insulin resistance may be induced in obese patients (Fasshauer *et al.*, 2001).

The association between the gene and protein levels was first determined in the Japanese population by Kondo *et al.* (2002). The mutation induces an amino acid alteration of an Ile164Thr within the globular domain of the protein. It was associated with a decrease in APM1 levels of the patients as well as an increased susceptibility towards T2D and various other features of the metabolic syndrome namely hyperlipidaemia, hypertension and atherosclerosis (Kondo *et al.*, 2002).

Further evidence for an association between this protein and BMI is presented by Stumvoll *et al.* (2002), via the investigation of a silent thymine to a guanine alteration within exon 2 of the APM1 gene. An association between this alteration, both in its homo- and heterozygotic forms, and an increase in insulin sensitivity, secondary to increased BMI was reported in this German cohort of patients, even upon division of the cohort into those with and those without a family history of T2D. It was concluded, however, that this association was only present within the group that did not have a family history of T2D. It has been hypothesised that the familial diabetic patients have a genetic background, which affects the phenotype so strongly that the detectable effect of this specific alteration is overpowered (Stumvoll *et al.*, 2002).

In the investigation by Vasseur *et al.*, 2002 it was determined that there were two promoter SNPs, namely C-11377G and G-11391A within the APM1 gene which were associated with increased risk to T2D within the French population. It was determined that both SNPs in the proximal region of the gene and non-synonymous alterations in the globular domains of the protein were associated with altered APM1 levels, thus contributing to diabetes risk. A risk haplotype within the promoter was also associated with both low APM1 levels and T2D. An earlier investigation into the Japanese diabetic population was undertaken by Hara *et al.* (2002) in which association was determined between the disease and an alteration termed T45G. The promoter SNPs may induce their effects by affecting the binding of transcription factors and disrupting regulatory elements whereas the missense alteration may affect or disrupt the formation of specific isoforms as well as the higher structure of this protein. Although association to T2D susceptibility was present for these alterations it was not possible to unequivocally determine linkage disequilibrium between any of these genetic alterations and disease susceptibility (Vasseur *et al.*, 2002). Therefore identification of the genetic variants responsible for the linkage between APM1 and T2D is essential. T45G is a thymine to a guanine alteration at nucleotide position 5,629 whereas C-11377G and G-11391A are in close proximity at nucleotide position 8,152 (cytosine to guanine) and at position 8,166 (guanine to adenine) respectively (GenBank accession number AF304467). The means by which these alterations affect disease susceptibility is not yet understood.

In an investigation in the German Caucasian population it was determined that a haplotype combination containing the wild-type allele at C-11377G and the variant allele at G-11391A was associated with a significantly elevated diabetes risk (Schwarz *et al.*, 2004). As previously stated a decrease in APM1 was associated with T2D. Thus if the

aforementioned haplotype combination within the promoter region of the APM1 gene could be directly associated with decreased expression of this protein it would be integral in defining the role of this protein in disease pathogenesis.

3.4 AIMS

The long-term objective of this research program is the eventual elucidation of the genetic aetiology of T2D susceptibility in order to understand disease pathogenesis. Due to the fact that the greatest increase in affected individuals will occur in the so-called developing countries it is necessary to investigate the genetic structures, which increase T2D susceptibility in these populations. In this investigation a meta-analysis will be undertaken by comparing the genetic risk factors at the CAPN10 and APM1 loci within a black South African and a Cuban cohort, and in turn comparing this to the risk genetic structures within these genes in a developed country specifically the German population. This investigation underscores the importance of utilising a population specific comparison group for the investigation of genetic susceptibility loci in T2D and possibly in all future metabolomic endeavours.

3.4.1 Specific aims

The specific aims of this project are as follows:

a) To screen the sample diabetic and control populations (both Southern African and Cuban) for the presence of the C-11377G and T45G SNPs within the APM1 gene as well as UCSNP-43, -44, -56 and -63 within the CAPN10 gene.

b) To compare the allelic and haplotypic frequencies of the aforementioned SNPs between the diabetic and control cohorts.

c) To determine if the risk profile of the black South African and Cuban cohorts are similar to that of a reported European population via a meta-analysis of these cohorts.

CHAPTER FOUR

MATERIALS AND METHODS

This research program has been approved by the Ethics Committee of the North-West University (Potchefstroom Campus). It has been approved under the title of, "Molecular analysis of non-insulin dependent diabetes mellitus (NIDDM) in the South African population" and has been assigned the approval number 02M08. The collection and analysis of the Cuban cohorts were approved by the Ethics Committee of the Carl-Gustav-Carus Medical Faculty, Technical University Dresden with the approval number EK16022000 (Schwarz, 2005).

All reagents utilised were, unless otherwise stated, analar grade products from Roche®[1]. Protocols supplied with the various kits were followed and deviations are noted in the text.

4.1 STUDY DESIGN

A case-control analysis strategy was followed in this investigation. Examination of specific SNPs within the CAPN10 and APM1 genes was undertaken within diabetic and control cohorts sampled from the black Southern African and Cuban populations. The genotypes, haplotypes and haplotype combinations generated from this molecular screening strategy were analysed via appropriate statistical measures e.g. χ^2 testing as well as calculation of odds ratios, in order to detect the T2D susceptibility associations that may exist in these cohorts.

4.1.1 Participant selection

Both T2D affected individuals and non-T2D affected individuals were sampled from the Cuban and black Southern African populations according to the criteria stipulated in Section 4.1.1.1. Neither of these two cohorts was stratified according to ethnicity but was rather grouped according to diabetic status as discussed in Section 5.1.1. Both patients and controls were collected with written informed consent preceding the collection of blood samples.

[1] Roche® is a registered trademark of Roche Molecular Biochemicals, Indianapolis, IN, USA.

4.1.1.1 Southern African cohort

The black South African control populations utilised during this investigation were collected from different outpatient clinics within the Gauteng and North West provinces of South Africa. The diabetic patients were collected from various diabetic clinics in the Gauteng area. All patients were diagnosed with the disorder via evaluation of the medical history of each individual. The diabetic phenotype was defined by at least two readings of a fasting plasma glucose level greater than 6.1 millimole.litre^{-1} (mmol.l^{-1}). All patients were selected to be over the age of 25 years, in order to decrease the probability of the inclusion of type 1 diabetics.

Selected control individuals were collected at hypertension outpatient clinics as these individuals were clinically well characterised and were not diagnosed with T2D. A random glucose measurement was taken of each control individual at the initial outset of the consultation. Control individuals were collected according to the following exclusion criteria:

- Individuals not undergoing treatment with any known anti-diabetic drug,
- Individuals with a random plasma glucose level of no greater than 6.1 mmol.l^{-1}

A group of 100 controls were collected in the Profiles of Obese Women with Insulin Resistance Syndrome (POWIRS) project conducted at the Metabolic Unit, North-West University (Potchefstroom Campus). Each patient underwent a five point OGTT. Samples were drawn subsequent to an overnight fast by the individuals, followed by the ingestion of 75 gram (g) glucose dissolved in 300 millilitres (ml) of water. Blood samples were collected at 30, 60, 90 and 120 minutes (min) post glucose challenge. Plasma and serum samples were centrifuged, collected and stored at –80 degrees Centigrade (°C) until analysis. Twenty ml of whole blood was collected in ethylenediamine tetra-acetic acid (EDTA) tubes for molecular genetic investigation.

4.1.1.2 Cuban cohort

The control and diabetic cohorts of the Cuban population were collected from various diabetic and outpatient clinics in Cuba. The diabetic and control individuals were enrolled according to the criteria stipulated in Section 4.1.1.1. No specific ethnic group was preferentially selected for in either cohort.

4.2 ISOLATION OF DNA

Isolation of DNA was achieved via the utilisation of a modified protocol of the NucleoSpin®[1] kit produced by Machery Nagel. Three ml of chilled C1 buffer was added to 3 ml of whole blood, followed by the addition of 9 ml of chilled sterile distilled water. The tube was repeatedly inverted until the solution became translucent. The sample underwent centrifugation at 1,300 x gravitational acceleration (x g) for 15 min at a temperature of 4°C, after which the supernatant was discarded.

The remaining leukocyte pellet was dissolved in 400 microlitres (µl) of phosphate buffered saline or PBS solution [4.3 millimolar (mM) disodium hydrogen phosphate (Na_2HPO_4), 1.4 mM potassium phosphate monobasic (KH_2PO_4) pH 7.4, 137 mM sodium chloride (NaCl), 2.7 mM potassium chloride (KCl)] via vortexing the solution for one min. The leukocytes were lysed by the addition of 400 µl buffer B3 and 50 µl of the proteinase K solution. The solution was in turn vortexed until the entire pellet had been dissolved after which the suspension was incubated at 70°C for 10 min to ensure complete lysis. Following incubation, 420 µl of 100% ethanol was added and the suspension was homogenised via vortexing.

A NucleoSpin® column was placed in a 2 ml collection tube where after 650 µl of the aforementioned solution was applied to the column. The column underwent centrifugation at 10,000 x g for 2 min in an Eppendorf®[2] 5810 R centrifuge utilising a fixed angle rotor. The column was removed and the flow through discarded, followed by one cycle of repetition of the previous step. The collection tube was replaced and 700 µl of wash buffer B5 was applied to the column. The matrix was washed by centrifugation at 10,000 x g for 1 min after which the flow through was discarded. An additional 300 µl of B5 buffer was added and the column was again washed for 2 min. DNA was eluted via the addition of 200 µl prewarmed elution buffer BE followed by a 15 min incubation step at 70°C. The elute was collected via centrifugation at 10,000 x g for 2 min. This step was repeated in order to collect 400 µl of eluted DNA solution.

The average yield of nucleic acid after isolation was determined via ultraviolet spectrophotometry. Absorbance of the sample was determined at both 260 and 280 nanometres (nm) to calculate both the nucleic acid and protein content, respectively.

[1] NucleoSpin® is a registered trademark of Machery Nagel GmbH & Co., Duren, Germany.
[2] Eppendorf® is a registered trademark of Eppendorf, Hamburg, Germany.

The concentration of the genomic DNA was estimated by the utilisation of Equation 4.1 described by Sambrook *et al.* (1989). Working dilutions consisting of a final concentration of 50 nanogram (ng).μl^{-1} were prepared for utilisation in the experimental phase and were stored at 4°C, whereas the stock solutions were maintained at -20°C.

Equation 4.1: Relationship of double stranded DNA concentration to ultraviolet absorbance

$$[\text{double stranded DNA}] = A_{260} \times (50\ \text{ng}.\mu l^{-1} \times \text{dilution factor})$$

A_{260} = Absorbance of sample at 260 nanometres.

4.3 STATISTICAL ANALYSES

The statistical methods utilised in the significance testing of the data generated within this investigation are presented in the following section. Comparison of the various cohorts was achieved utilising standard contingency table analyses, and the level of significant difference was calculated utilising χ^2 analysis. The risk of the specific genetic factors investigated was calculated utilising odds ratios. Determination of global effects was achieved via a meta-analysis of the German, black Southern African and Cuban cohorts discussed in Section 5.3.5.

4.3.1 Calculation of significance level

Due to multiple testing of several SNPs at various loci, in the same cohorts in this investigation, the significance level had to be adjusted. To achieve this, the Bonferroni method (Bland and Altman, 1995) as described in Equation 4.2 was utilised. The adjusted significance level is calculated utilising two values namely the significance level that would be accepted if only a single sample was investigated (α), and the number of independent samples (κ). The frequency of the correct acceptance of a significant result was generated by subtracting the original significance level from 1. This was raised to the power of 1 over the number of independent samples, followed by subtraction of this result from 1. This indicates the adjusted significance level.

Equation 4.2: Calculation of the adjusted significance level

$$\alpha' = 1-(1-\alpha)^{1/\kappa}$$

α = significance level; α' = adjusted significance level; κ = number of independent samples.

4.3.2 Hardy-Weinberg equilibrium

The allele frequencies (p = frequency of the 1 allele; q = frequency of the 2 allele) of the various alterations were determined utilising the equations as described in Equation 4.3 (Hardy, 1908). This was followed by χ^2 testing to ensure that the populations screened were in Hardy-Weinberg (HW) equilibrium with respect to the specific alteration.

Equation 4.3: Calculation of the allele frequencies for alleles 1 and 2

Allele 1:
$$p = \frac{[\text{number of homozygotes} + \frac{1}{2}(\text{number of heterozygotes})]}{\text{total number of individuals}}$$

Allele 2:
$$q = 1-p$$

p = frequency of the 1 allele; q = frequency of the 2 allele.

Throughout this investigation p always refers to the frequency of the 1 allele. In the above equation the number of homozygotes thus refers to the number of 1,1 homozygotes. The letter q refers to the frequency of the 2 allele.

At all the loci to be investigated there are only two alleles described. Generally the non-risk allele was designated as 1 and the "at risk" allele as 2. It was possible to discern three genotypes namely the homozygotes 1,1 and 2,2 as well as the heterozygote 1,2.

By comparison of the observed frequencies of the alleles to the expected frequencies it was possible to determine whether the population was in HW equilibrium. Expected ratios of genotypes were determined as functions of allele frequencies as described by Hardy in 1908. The probability of being homozygous (f [1,1] or f [2,2]) for a specific allele at a locus was equal to the frequency of that allele in the population squared (p^2 or q^2). The probability of heterozygosity (f [1,2]) was equal to twice the product of the frequencies of

the 1 and 2 alleles ($2pq$). The mathematical equations for these calculations are depicted in Table 4.1.

Table 4.1: Calculation of the expected genotype frequency utilising allele frequencies

$$f[1,1] = p \times p = p^2$$
$$f[1,2] = 2pq$$
$$f[2,2] = q \times q = q^2$$

f [1,1] = probability of a 1,1 homozygote; f [1,2] = probability of a heterozygote; f [2,2] = probability of a 2,2 homozygote; p = frequency of the 1 allele; q = frequency of the 2 allele.

The significance of the difference between the observed and expected numbers was determined via χ^2 testing of the resultant frequencies. This was achieved by squaring the difference between the observed and expected number ($[O-E]^2$). The dividend of this result and the expected number as depicted in Equation 4.4 was calculated for each specific genotype, and the sum of these calculations determined. If the sum did not exceed a critical value of $\chi^2 = 9.21$ ($\alpha' = 0.01$, degrees of freedom (df) = 2) the population was deemed to be in HW equilibrium and allowed to undergo further statistical analyses.

Equation 4.4: Chi square test

$$\chi^2 = \frac{(O-E)^2}{E}$$

χ^2 = chi square value; O = observed number of a specific genotype; E = expected number of a specific genotype as determined via calculation.

The determination of the significance level is discussed in Section 5.3.2 whereas the calculation of the df was achieved by utilising the numbers of rows (r) and columns (k) utilised in the contingency table. The exact equation used to calculate the df is presented in Equation 4.5.

Equation 4.5: Calculation of degrees of freedom

$$df = (r-1)(k-1)$$

df = degrees of freedom; r = number of rows within the contingency table; k = number of columns within the contingency table.

4.3.3 Chi square analysis

During this investigation it was necessary to utilise various statistical analyses in order to determine the significance of any variation established. The comparison of the allele and haplotype frequencies between the patient and control populations was achieved utilising contingency table analysis as is generally applied to categorical data (Samuels *et al.*, 1989).

In determining the statistical significance of the variation in allele frequencies between the various populations it was necessary to utilise the χ^2 test to determine which of the null hypothesis (H_0) or the alternative hypothesis (H_A), were statistically more valid. The definition of both these hypotheses, as utilised in this investigation, are presented in Table 4.2.

Table 4.2: Definition of the null and alternative hypotheses for the statistical analysis of genotype frequencies between the patient and control cohorts

H_0	The genotype frequency of a specific alteration is the same in the patient cohort under investigation as in the cohort of control individuals
H_A	The genotype frequency of a specific alteration is different in the patient cohort under investigation as compared to the cohort of control individuals

H_0 = null hypothesis; H_A = alternative hypothesis.

The equation depicted in Equation 4.4 was utilised to determine the χ^2 value with the observed number representing the actual genotype number within the patient population investigated while the expected values were determined from the control population. This was followed by comparison of the χ^2 value to that of the critical χ^2 value. If a calculated χ^2 value was lower than the critical value, increased relevance was attributed to the null hypothesis. However, if the aforementioned value was higher than the critical value, evidence was generated in favour of the alternative hypothesis.

Following determination of the various combinations as discussed in Section 4.4.6 and 4.4.11, the haplotype frequencies were calculated via a similar method as the genotype frequencies as depicted in Table 4.1. In this calculation the dividend of the number of a specific haplotype and the total number of haplotypes was computed. These frequencies were utilised in the comparison of specific haplotype combinations present within the control populations. Haplotype frequencies were determined for both the patient cohort (observed number) and the control cohort (expected number) following molecular

determination of the alleles present at each locus. These values were utilised in a χ^2 test as described in Equation 4.4 with the H_0 and H_A as defined in Table 4.3.

Table 4.3: Definition of the null and alternative hypotheses for the statistical analysis of haplotype frequencies between the patient and control cohorts

H_0	The haplotype frequency for a specific gene is the same in the patient cohort under investigation as in the cohort of control individuals
H_A	The haplotype frequency for a specific gene is different in the patient cohort under investigation as in the cohort of control individuals

H_0 = null hypothesis; H_A = alternative hypothesis.

4.3.4 Odds ratio determination

Although determining the association between a specific genetic structure and the disease is important, it is still necessary to measure the strength of the effect that is present. This was determined via the calculation of the OR as presented in Equation 4.6 (Bland and Altman, 2000). It was determined by firstly calculating a numeric value for the evidence validating the association between a specific genetic structure and the disease followed by determining the ratio of that value in comparison to a second value representing the evidence against this aforementioned association.

To achieve this it was first necessary to determine the product of the number of patients harbouring the genetic structure (n_1), and the number of control individuals in which it was absent (n_4). Secondly it was necessary to calculate a similar product to describe the number of control individuals in which it was present (n_2) and the number of patients which did not harbour it (n_3). The dividend of these two values was the odds ratio of the specific alteration. The implications of the calculated OR value are discussed in Section 5.3.4.

Equation 4.6: Odds ratio determination

Genetic factor	Patients	Control	$OR = \frac{n_1 n_4}{n_2 n_3}$
Present	n_1	n_2	
Absent	n_3	n_4	

OR = odds ratio; n_1 = patients harbouring the genetic factor; n_2 = controls harbouring the genetic factor; n_3 = patients not harbouring the genetic factor; n_4 = controls not harbouring the genetic factor.

This method may seem rather simplistic but it is extremely useful in determining the strength of the association as discussed in Section 5.3.4. In the utilisation of this form of analyses it was however necessary to define the level of precision of the OR value. This was achieved via the calculation of the CI. Determination of the CI was slightly more

complicated than the calculation of the odds ratio itself. It was decided that Woolf's method of CI calculation (Woolf, 1955) would be utilised as presented in Equation 4.7. For this calculation it was necessary to again divide the cohort into the four groups (n_1-n_4) as listed in Equation 4.6. For each group the inverse of the sample size was determined after which the sum of all four groups was calculated.

The larger the values of n the smaller the end result, ultimately leading to a smaller confidence interval. The square root of this value was determined, and in turn was multiplied by 1.96 as a confidence level of 95% was required. The natural logarithm of the OR was calculated and the aforementioned value was added or subtracted in order to determine the upper and lower limit respectively. The resultant values were logarithmic in nature and to define the 95% CI in terms of the OR it was necessary to raise the natural logarithmic base (e) to the power of the defined upper and lower limits. The equation utilised to calculate the 95% CI is presented in Equation 4.7.

Equation 4.7: Determination of confidence interval

$$95\%\ CI = \exp\left(\ln OR \pm 1.96\left(\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}}\right)\right)$$

CI = confidence interval; OR = Odds ratio; n_1 = patients harbouring the genetic factor; n_2 = controls harbouring the genetic factor; n_3 = patients not harbouring the genetic factor; n_4 = controls not harbouring the genetic factor.

4.3.5 Meta-analysis

Determination of global associations of specific genotypes was calculated via the utilisation of a meta-analysis of the black Southern African, Cuban and a reported German population. These investigations were combined under two separate models, namely that of fixed effects and random effects. The utilisation of these two models is discussed in Section 5.3.5. Furthermore within each model, five possible types of genotypic association were investigated as depicted in Table 4.4. A two row by two column contingency table was constructed with a cohort structure similar to that depicted in Equation 4.6 for each population group investigated. For each type of genotypic association, only the genotypes as presented in Table 4.4 were utilised i.e. for the recessive model the individuals harbouring the 1,2 and 2,2 genotypes were pooled and compared to the 1,1 genotype. Via this method it was possible to determine the OR value for a specific set of cohorts termed OR_i as indicated in Equation 4.8a. This describes the risk of a specific genotype within the cohort under investigation and is derived in a similar manner as the OR presented in

Equation 4.6. All calculations were performed by the utilisation of the Microsoft™[1] Excel software.

Table 4.4: Genetic association models investigated via meta-analysis

Model	Assumption	Model	Assumption
Recessive	1,1 vs 1,2 and 2,2	Additive	1,1 vs 1,2
			1,2 vs 2,2
Dominant	1,1 and 1,2 vs 2,2		1,1 vs 2,2

1,1 = homozygote for the 1 allele; 1,2 = heterozygote; 2,2 = homozygote for the 2 allele.

4.3.5.1 Meta-analysis under fixed effects

Under fixed effects, Wolfe's method of meta-analysis was utilised in order to calculate an overall OR score (OR_{wolfe}) for the three groups investigated. The natural logarithm of the value calculated via Equation 4.6 was utilised as OR_i. Each group investigated was weighted according to the cohort size as well as the level of internal variance to determine its importance within the meta-analysis. This weighted value is termed w_i and was calculated by determining the inverse of the internal variance (var_i) within a population as depicted in Equation 4.8b. The var_i was calculated, as presented in Equation 4.8c, as the sum of the inverses of the various cohorts as defined in Equation 4.6. By determining the dividend of the sum of the weighted natural logarithm of the OR_i for each population group investigated, and the sum of w_i assigned to each population group, it was possible to determine the natural logarithm of the OR_{wolfe}. However as previously discussed in Section 5.3.4 in order to determine the relevance of an OR value, a 95% CI must be generated. As outlined in Equation 4.8d, the 95% CI was calculated utilising the natural logarithm of the OR_{wolfe} value to which a specific value is added and subtracted. This value is the inverse of the square root of the sum of the w_i multiplied by a constant value of 1.96. Similar deductions were made as discussed in Section 5.3.4 with regard to the 95% CI.

In this analysis it was necessary to determine a measure of heterogeneity (Q_{wolfe}) in order to calculate the OR of each genotype under random effects. This value was defined by the application of Equation 4.8e, where the sum of the weighted difference of the natural logarithm of the OR_i of each population group and the natural logarithm of the OR_{wolfe} is determined to define Q_{wolfe}.

[1] Microsoft® is a registered trademark of the Microsoft Corporation, Seattle, WA, USA.

Equation 4.8: Statistical determination of global odds ratio and 95% confidence interval utilising Wolfe's method for the fixed effects meta-analysis model

Equation 4.8a	$\ln OR_{wolfe} = \frac{\sum w_i \ln OR_i}{\sum w_i}$
Equation 4.8b	$w_i = \frac{1}{var_i}$
Equation 4.8c	$var_i = \frac{1}{n_{1i}} + \frac{1}{n_{2i}} + \frac{1}{n_{3i}} + \frac{1}{n_{4i}}$
Equation 4.8d	$\ln OR_{wolfe\text{-}95\%CI} = \ln OR_{wolfe} \pm 1.96 \frac{1}{\sqrt{\sum w_i}}$
Equation 4.8e	$Q_{wolfe} = \sum w_i (\ln OR_i - \ln OR_{wolfe})^2$

OR_i = odds ratio calculated to describe the risk within a specific population group; OR_{wolfe} = global odds ratio determined via Wolfe's method under the fixed effects model; $OR_{wolfe\text{-}95\%CI}$ = 95% confidence interval calculated for OR_{wolfe}; w_i = weight given to a specific population in a meta-analysis under the fixed effects model; var_i = internal variance within a specific population group; Q_{wolfe} = heterogeneity statistic determined via Wolfe's method under the fixed effects model; n_{1i} = patients harbouring the genetic factor; n_{2i} = controls harbouring the genetic factor; n_{3i} = patients not harbouring the genetic factor; n_{4i} = controls not harbouring the genetic factor.

4.3.5.2 Meta-analysis under random effects

Under random effects, both intra and inter population variance were taken into account by the use of the DerSimonian and Laird method. This was achieved by including the measure of inter population variability (τ) in the equation for the derivation of the weighted value according to the DerSimonian and Laird method (w_{DL}) as presented in Equation 4.9c.

This value is defined as the dividend of the difference between Q_{wolfe} and one less than the number of population groups investigated (k), and a specified divisor. This divisor is calculated by determining the difference of the sum of w_i and the dividend of the sum of squared w_i and the sum of w_i. The weighted value determined via this method (w_{DL}) is similar to that calculated in Equation 4.8, however the inter population variability is included in Equation 4.9b by the addition of τ.

The natural logarithm of OR determined via this method (OR_{DL}) is calculated similarly to Wolfe's method when Equations 4.8a and 4.9a are compared, the difference being in the calculations of the w_{DL}. Definition of 95% CI via the DerSimonian and Laird method is also

calculated in a similar manner as depicted in Equation 4.9d, than that calculated via Wolfe's method as depicted in Equation 4.8d.

Equation 4.9: Statistical determination of global odds ratio and 95% confidence interval utilising the DerSimonian and Laird method for the random effects meta-analysis model

Equation 4.9a	$\ln OR_{DL} = \frac{\sum w_{DL} \ln OR_i}{\sum w_{DL}}$
Equation 4.9b	$w_{DL} = \frac{1}{var_i + \tau}$
Equation 4.9c	$\tau = \frac{Q_{wolfe} - (k-1)}{\sum w_i - \frac{\sum (w_i)^2}{\sum w_i}}$
Equation 4.9d	$\ln OR_{DL\text{-}95\%CI} = \ln OR_{DL} \pm 1.96 \frac{1}{\sqrt{\sum w_{DL}}}$

OR_{DL} = global odds ratio calculated via the DerSimonian and Laird method under the random effects model; $OR_{DL\text{-}95\%CI}$ = 95% confidence interval calculated for OR_{DL}; w_{DL} = weight given to a specific population in a meta-analysis under the random effects model which includes the effects of both inter and intra population variability; var_i = internal variance within a specific population group; Q_{wolfe} = heterogeneity statistic determined via Wolfe's method under the fixed effects model; k = number of population groups investigated; τ = measure of inter population variability.

4.4 REAL TIME POLYMERASE CHAIN REACTION AND MELTING CURVE ANALYSIS

All genetic analyses were performed at the Carl-Gustav-Carus Medical Faculty, Technical University Dresden in Dresden, Germany. The LightCycler™ (LC)-FastStart™[1] Plus DNA Master Hybridisation Probes kit was utilised in the real-time PCR strategy for the detection of the various alterations discussed. This kit has been completely optimised by the manufacturer for allele detection utilising hybridisation probe chemistry. It was therefore unnecessary to alter the magnesium chloride concentration during optimisation hence simplifying the entire process. DNA amplification via the action of a thermostable polymerase as described by Mullis *et al.* (1986, 1987) is the basic principle from which this technique was derived.

The kit contained the LC™-FastStart™ polymerase which is a "hot-start" enzyme, thus its action is prevented via treatment with a heat-labile enzyme-specific antibody as described by Kellog *et al.* (1994). This prevents the occurrence of any secondary amplification preceding the initial denaturation step, which may interfere with the fluorescent signal and prevent accurate downstream analysis.

[1] FastStart™ is a trademark of Roche Molecular Biochemicals, Indianapolis, IN, USA.

Furthermore, the kit also contained the FastStart™ Plus Reaction Mix Hybridisation Probes 10 x reaction buffer which consisted of a deoxynucleotide triphosphate (dNTP) mixture wherein, deoxythymidine-5'-triphosphate (dTTP) had been replaced with deoxyuracil-5'-triphosphate (dUTP). Carry over contamination can therefore be prevented by treatment with a heat labile uracil-DNA glycosylase (Roche Molecular Biochemicals, Hamburg, Germany). However, this was not required in this investigation.

The LC™ reaction utilised in the screening of these variants consisted of the following components namely 10 picomole (pmol) of each of the forward and reverse primers (presented in Table 4.5), 3 pmol of the LC640 probe (presented in Table 4.7) and 1.5 pmol of the sensor probe (presented in Table 4.7).

Preceding the addition of the enzyme to the reaction mixture, one vial of both the enzyme and the reaction mix underwent centrifugation followed by the transfer of 60 µl of the reaction mix to the tube containing the enzyme. The solution was homogenised and relabelled to prevent accidental misuse, followed by the addition of 2 µl to each reaction. The sequence of the primer sets utilised in the amplification of the various regions within the APM1 and CAPN10 genes are depicted in Table 4.5.

Table 4.5: Primers utilised in the amplification of regions harbouring specific single nucleotide polymorphisms within the calpain 10 and adiponectin genes

	UCSNP	*Primer*	*Primer sequence*[1]
Calpain 10	43 + 44	UCSNP-43/44 F1	**F: 5'**-tgt cgg cac acc gga tgc-**3'**
		UCSNP-43/44 R1	**R: 5'**-ggt ctg tag cac ccc aaa tcg-**3'**
	56	UCSNP-56 F1	**F: 5'**-ggc ctc agg cac act gta g-**3'**
		UCSNP-56 R1	**R: 5'**-tct tcc tgc ctc gca cta g-**3'**
	63	UCSNP-63 F1	**F: 5'**-cac tcg gtc aga gcc cta gc-**3'**
		UCSNP-63 R1	**R: 5'**-ggt gcc tga agg ttc cac tc-**3'**
Adiponectin	***SNP***	***Primer***	***Primer sequence***[2]
	C-11377G G-11391A	ACR1F	**F: 5'**-act tgc cct gcc tct gtc tg-**3'**
		ACR1R	**R: 5'**-gcc tgg aga act gga agc tg-**3'**
	T45G	ACR3F	**F: 5'**-aga aag cag ctc cta gaa gt-**3'**
		ACR3R	**R: 5'**-ggc acc atc tac act cat cc-**3'**

[1]Primer sequences as described by Görgens *et al.* (2003); [2]Primer sequences as described by Schwarz *et al.* (2004); UCSNP = University of Chicago single nucleotide polymorphism; SNP = single nucleotide polymorphism; F and R represent forward and reverse primers respectively.

In all the reactions utilised in the detection of these alterations a master mixture of the various components was prepared according to the concentrations previously discussed. Eighteen μl of this solution was aliquoted into a LightCycler™ capillary tube to which 100 ng of patient genomic DNA was added. The capillary tube was capped and centrifuged to ensure the reaction volume filled the entire capillary. The reaction was allowed to undergo amplification in the LightCycler™ instrument and detection of the resulting product was achieved via the utilisation of specific LightCycler™ master hybridisation probes. Amplification was achieved in a LightCycler™ Real Time PCR Machine via the thermal cycling conditions presented in Table 4.6. In the case of UCSNP-43, -44, -56 and -63 initial denaturation was followed by 40 cycles of the aforementioned program, whereas the SNPs within the APM1 gene were subjected to 35 cycles of this program.

Table 4.6: Thermal cycling conditions utilised in the real time polymerase chain reaction

Stage	Molecular reaction	Temperature	Time
Amplification	Initial denaturation	95°C	10 minutes
	denaturation	95°C	10 seconds
	annealing	T_a	7 seconds
	elongation	72°C	9 seconds
Melting curve	annealing	Thermal ramp to 30°C	variable
	detection of denaturation	Thermal ramp to 94°C	0.5°C.s^{-1}
	final denaturation	94°C	1 second

T_a = annealing temperature; s = second; °C = degrees Centigrade.

Melting curve analysis was achieved via the conditions presented in Table 4.6. The temperature was decreased to 30°C to ensure that most of the probe was annealed to the template. The temperature was increased by 0.5°C.second.$(s)^{-1}$ up to a final temperature of 94°C, so as to detect the temperature at which denaturation of the probe and template occurred. Fluorescence was detected continuously throughout this process and was plotted graphically as fluorescence versus temperature by the LightCycler™ software. The melting peak of each SNP was detected by determining the negative derivative of the fluorescence and plotting it against temperature.

Detection of PCR product and determination of the melting curve was achieved by fluorescence detection utilising specific hybridisation probes. The various probe sets employed in the screening of UCSNP-43, -44, -56 and -63 as well as C-11377G, T45G and G-11391A within the APM1 gene are presented in Table 4.7. The level of fluorescence

was determined once per cycle of PCR after the elongation step and continuously throughout the melting curve analysis.

Table 4.7: Sequences of the hybridisation probes utilised in detection of the various SNPs within the calpain 10 and adiponectin gene

Calpain 10		
UCSNP	***Probe***	***Probe sequence[1]***
43	43Gwt	**5'-** tca cct tca aac gcc tta ctt c-**X-3'**
	43+44LC640	**5'-** LC640-gcg tga gcg ccc tgc agt cct ct-**p-3'**
44	44 wtFL	**5'-** gcc tta ctt cac agc a-**X-3'**
	43+44LC640	**5'-** LC640-gcg tga gcg ccc tgc agt cct ct-**p-3'**
56	56mtFL	**5'-**cct tcc cct act gtc ctc ttc cag-**X-3'**
	56LC640	**5'-** LC640-gga cgt ggc cct tct ctc ccc t-**p-3'**
63	63Cmt	**5'-** ccc cct cgc tcc acc cc-**X-3'**
	63LC640	**5'-** LC640-tca ggc cct ggc ccc cct tg-**p-3'**
Adiponectin		
SNP	***Probe***	***Probe sequence[2]***
C-11377G	SNP1anchor	**5'-**LC640-aca tga gcg tgc caa gaa agt cca agg tgt tg **-p-3'**
	SNP1sensor	**5'-**ctc aga tcc tgc cct tca aaa ac-**X-3'**
T45G	SNP3newss	**5'-**LC640-ttc ctg gtc atg ccc g **-p-3'**
	SNP3newar	**5'-**agg act ccg ggc cct tga gtc gt-**X-3'**
G-11391A	AnchorAPM10-1	**5'-**LC640-gca agc cac aca ttc tga tga att aaa tta cga ccc-**p-3'**
	SensorGsnp10	**5'-**gca gga tct gag ccg gtt ct-**X-3'**

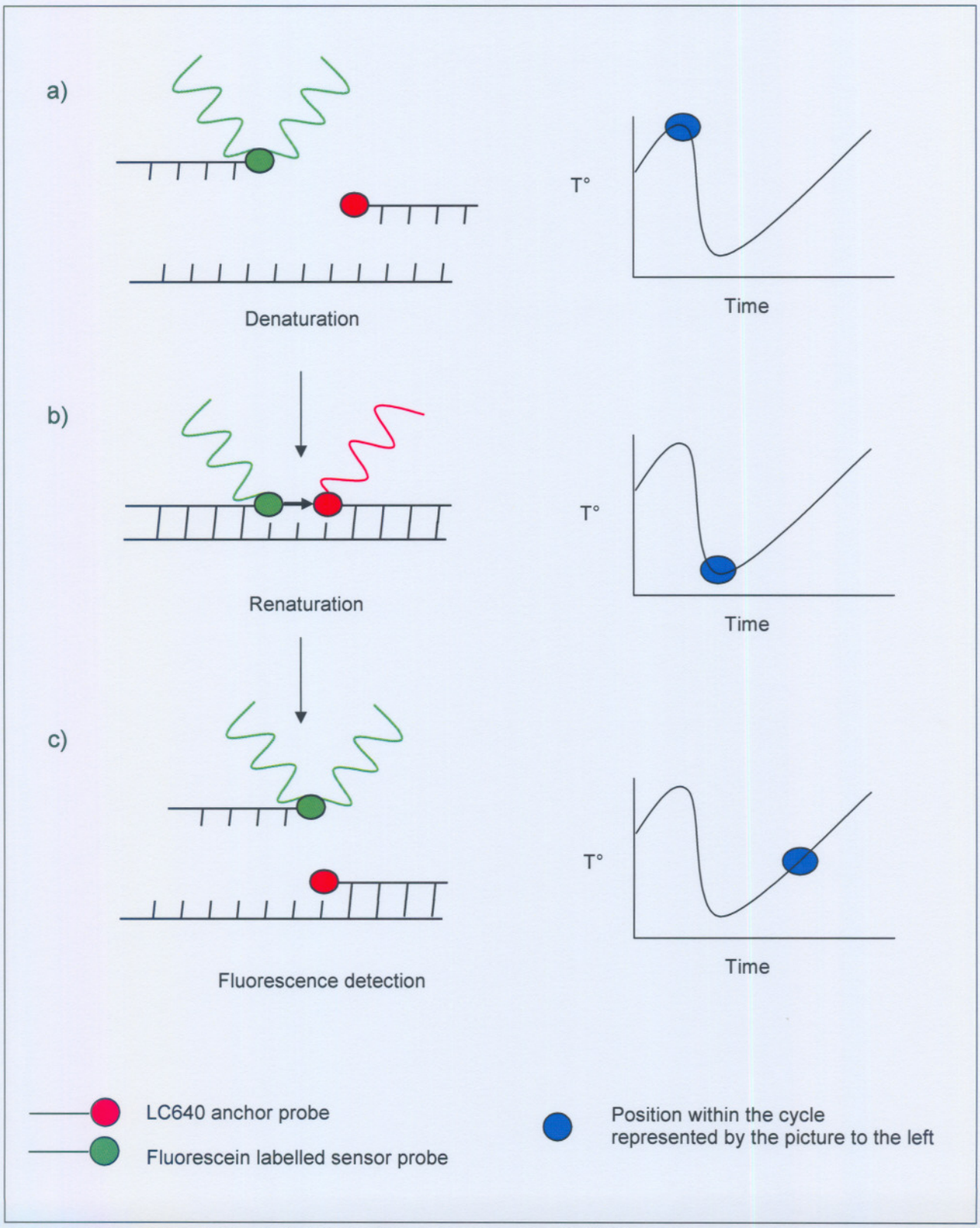
[1]Probe sequences as described by Görgens *et al.* (2003); [2]The primer sequences were described by Schwarz *et al.* (2004); SNP = single nucleotide polymorphism; X = fluorescein; LC640 = LightCycler™ 640 red fluorophore; p = phosphorylated 3' end.

The utilisation of hybridisation probes ensured specific recognition of the regions of interest. As depicted in Figure 4.1, each probe had a fluorescent probe attached to either the 5'- or 3'-end. One fluorophore was excited by the application of a specific wavelength as presented in Figure 4.1a. The excitation spectrum of the second fluorophore had to overlap the emission spectrum of the first fluorophore. For this reason fluorescein was utilised. Emission at a wavelength of 530 nm was obtained which excited the LC640 probe. Due to the overlap of the emission spectra of these two probes, energy was transferred to the LC™ probe and caused it to fluoresce. When the two probes were in close proximity as in Figure 4.1b, it allowed for the detection of a signal different from that expected if a single fluorophore emitted the signal.

During the melting curve analysis the temperature of the reaction mixture was slowly increased. It was possible to detect the point at which the probes denatured from the template by a decrease in fluorescence at 640 nm as outlined in Figure 4.1c. If the probes

were specific for the wild type allele and the template harboured the mutant allele the decrease in fluorescence would occur at a lower temperature due to mismatching of the probe. The mutant and wild type alleles were thus differentiated in this manner.

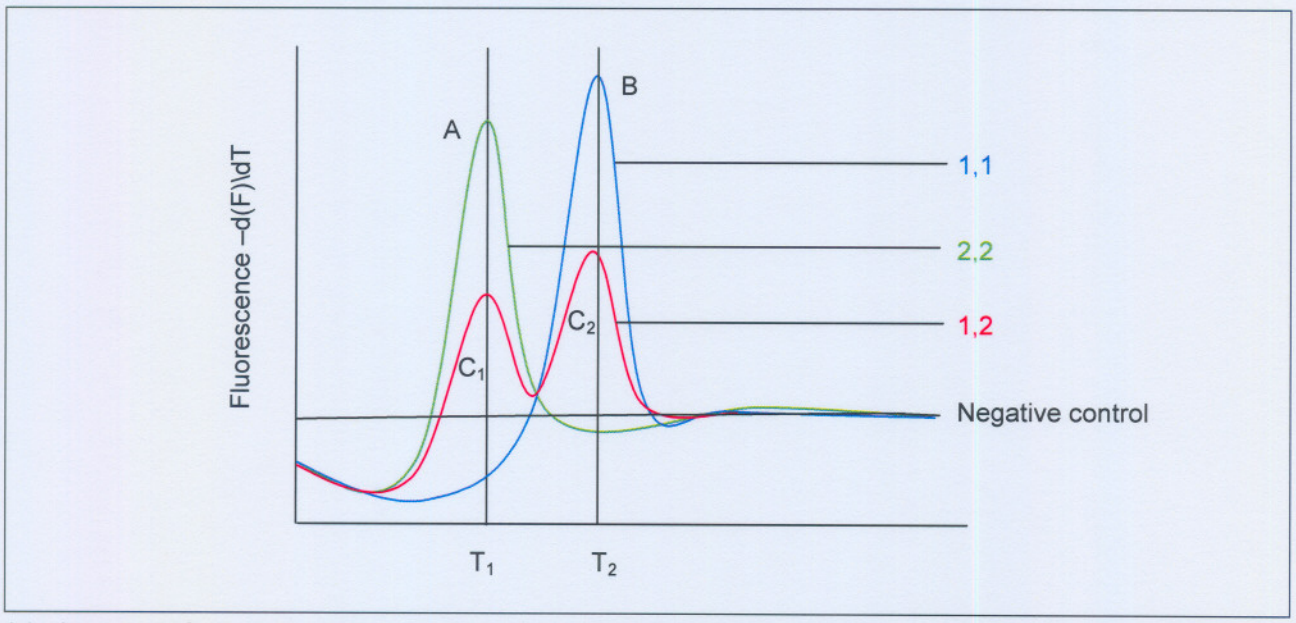
Figure 4.1: Diagrammatic representation of fluorescence resonance energy transfer (FRET) technology



T° = temperature, LC640 = LightCycler fluorophore which emits at a wavelength of 640 nanometres.

After detection the signal intensity was graphically compared to time by the LightCycler™ software. The melting peak was generated via the calculation of the negative differential of the fluorescence level versus time following detection (y = negative differential of the fluorescence level (-d(F)) with respect to time (dT)). The different peaks generated are depicted in Graph 4.1. Exact complementarity of the probe to the template resulted in a single peak at the melting temperature of the probe as indicated at curve B in Graph 4.1. Mispairing, due to the presence of the mutation caused denaturation of the probe at a lower temperature thus producing a peak as that present at A in Graph 4.1. In the case of a heterozygotic individual there was two peaks produced however the amplitude was half that of a peak detected for a homozygote as depicted by peaks C_1 and C_2 in Graph 4.1.

Graph 4.1: Diagrammatic representation of the differential graph of probe fluorescence versus time



1,1 = homozygote for the 1 allele indicated by the blue line; 1,2 = heterozygote indicated by the red line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by black line wherein DNA was replaced by deionised water; T_1 = temperature at which the 2,2 homozygote peaks; T_2 = temperature at which the 1,1 homozygote peaks.

4.4.1 Detection of single nucleotide polymorphisms in the calpain 10 gene

Due to the reported association of the SNPs discussed in Section 4.4.1.1 to 4.4.1.4 with T2D (Horikawa *et al.*, 2000) they were screened in the cohorts included in this study. These alterations were all detected by real time PCR. The location of each SNP within the genomic sequence of human chromosome 2q37.3 (GenBank accession number AF158748) is presented in the following sections as well as the annealing sites of the relevant probes and primers. If the published sequence was not exactly complementary to the sequence of the sensor probe, e.g. due to the variant allele being present in the published sequence as opposed to the wild type allele recognised by the probe, the sequence was altered accordingly. This was undertaken to ensure that the probe

sequences listed in Table 4.7 could be compared to the various sequences presented. All alterations made in the sequence are indicated within the text preceding the partial genomic DNA sequence. All sequence presented in the following sections, is indicated in the forward direction.

4.4.2 UCSNP-43 in the calpain 10 gene

Both UCSNP-43 and -44 are present within intron three of the CAPN10 gene and were therefore both amplified utilising the primer set UCSNP-43/44 F1 and UCSNP-43/44 R1. However, detection of each was achieved utilising different probe sets. The position and nature of UCSNP-43 (allele 1 = guanine, allele 2 = adenine) is indicated in Table 4.8. This alteration was screened via FRET technology utilising the LightCycler™ apparatus.

Table 4.8: Allele composition of UCSNP-43 and -44 within the calpain 10 gene

UCSNP	nt change	Position	UCSNP	nt change	Position
43	G>A	22,762	44	T>C	22,751

UCSNP = University of Chicago single nucleotide polymorphism; Nucleotide numbering according to Horikawa *et al.* (2000).

The primer sequences represented in Table 4.5 were employed in the amplification of a segment of intron three. The primer and hybridisation probe annealing sites for the amplification of UCSNP-43 are indicated on the partial sequence of intron three presented in Table 4.9. The sensor probe has been designed to specifically recognise the more prevalent allele and the sequence has been altered accordingly i.e. at position 22762 an adenine has been replaced by a guanine.

Table 4.9: Partial sequence of intron three of the calpain 10 gene from nucleotide 22561 to 22980

Nucleotide number	DNA Sequence: UCSNP-44 and -43
22561	tggtgacatc agtgcccagt gagcccttcc atcccaaggg ctgttttagg aaaagcaggg
22621	ttggagcttg agagccaagg gatgtgggca tccatagctt ccacgcctcc tgccctgctc
22681	ctgtgcccac accggatgcc agagagtttc tgtgtgtggg cagaggactg cagggcgctc
22741	acgcttgctg tgaagtaagg cgtttgaagg tgaggctaag ccttgacttg gtgaggatga
22801	ggaagaaggc agaggggagt aaagaggtgg gattgaggca gcggttggac gatttggggt
22861	gctacagacc atgggaatca gagagggggc catgctcaat gccagaggct cactcccatg
22921	gtgattgtgt cccctagggt ccatgggtcc tacgagcacc tgtgggccgg gcaggtggcg

The forward primer sequence is indicated by the double underlined text (vvv) whereas the position of the reverse primer sequence is indicated by the underlined text (www). The anchor probe sequence is indicated by the red text (xxx). The sensor probe 43wtFL specific for UCSNP-43 is indicated by the blue text (yyy) whereas the sensor probe 44wtFL specific for UCSNP-44 is indicated by the text highlighted in yellow (zzz). A block (□) indicates the position of UCSNP-44 and a circle (○) indicates that of UCSNP-43. This sequence has been retrieved from GenBank accession number AF158748. Nucleotide numbering is according to Horikawa *et al.* (2000).

4.4.3 UCSNP-44 in the calpain 10 gene

As presented in Section 4.3.1.1, UCSNP-44 (allele 1 = thymine, allele 2 = cytosine) is in intron three of the CAPN10 gene and was amplified utilising the primers termed UCSNP-43/44 F1 and UCSNP-43/44 R1. The nucleotide position as well as the different alleles present at this locus, are presented in Table 4.8. The annealing sites of the primers and probes utilised in the determination of the genotype at this locus are presented in Table 4.9. The sequence has been appropriately altered at position 22751 i.e. a cytosine has been replaced by a thymine, to ensure the exact complementary sequence to the wild type specific sensor probe is presented.

4.4.4 UCSNP-56 in the calpain 10 gene

Although both loci are within intron six, the utilisation of UCSNP-56 (allele 1 = adenine, allele 2 = guanine) in favour of UCSNP-19 is discussed in Section 5.3.1.3. The recognition sites of the primers and hybridisation probes utilised in the detection of UCSNP-56 are represented in Table 4.10. The sensor probe is specific for the wild type SNP, therefore the sequence presented has been altered accordingly i.e. an adenine replaces a guanine at nt position 23325.

Table 4.10: Partial sequence of intron six of the calpain 10 gene from nucleotide 23101 to 23460

Nucleotide number	DNA Sequence: UCSNP-56
23101	ctccacctga aggaccagtg tctgatcagc tgctgcgtgc tcagccccag agcaggtgag
23161	gcacgtggcc aacatgggag ggctgcagcc agcgtgcccc ccactgccag gcctcaggca
23221	cactgtagct ttttatgtga ctggctacac agccctgtca ggactaagtg ggaagaagta
23281	agcttgttct caagggtggt gtcctcagtt tgtgaccttc ccct□ctgtc ctcttccaga
23341	gggacgtggc ccttctctcc cctgaccagt cctttccact agtgcgaggc aggaagaggt
23401	ggcaccgagt caaagcccac tgtctgtgcc atccctggcc cagctggcaa cctggcaaaa

The forward primer sequence is indicated by the double underlined text (vvv) and the position of the reverse primer is indicated by the underlined text (www). The anchor probe sequence is represented by the text in red (xxx). The sensor probe is indicated by the text in blue (yyy). A block (□) indicates the position of UCSNP-56. The sequence has been retrieved from GenBank accession number AF158748. Nucleotide numbering is according to Horikawa *et al.* (2000).

4.4.5 UCSNP-63 in the calpain 10 gene

This polymorphism is present in intron 13 and is a transition alteration of a cytosine (allele 1) to a thymine (allele 2) at nucleotide 34288 (Horikawa *et al.*, 2000). It was determined via real time PCR and melting curve analysis, utilising the primer set UCSNP-63 F1 and UCSNP-63 R1 presented in Table 4.5. The probes were designed to

recognise the wild type allele. Primer and probe annealing sites are depicted in the sequence presented in Table 4.11.

Table 4.11: Partial sequence of intron thirteen of the calpain 10 gene from nucleotide 34141 to 34440

Nucleotide number	DNA Sequence: UCSNP-63
34141	gtccccagct ggctgggcct gcagcccccc cctgtgcccc gagctggccg ggcccgcagc
34201	ccactccctg gtcactggat gttgctgaca cttcactcgg tcagagccct agcaccaag
34261	gggggccagg gcctgacggg ggtggagcga ggggggtggc cgcgtctgtg caggctcaag
34321	aagcttccta agaggctgga gagtggaacc ttcaggcacc acgcactgcc tcctccctgc
34381	ccacggtcct gggtttctcc agatggggcc ttggccttgg ctaggtgttg atcaggagct

The forward primer sequence is indicated by the double underlined text (vvv) and the position of the reverse primer is indicated by the underlined text (www). The anchor probe sequence is represented by the text in red (xxx). The sensor probe is indicated by the text in blue (yyy). The block (□) indicates the position of UCSNP-63. The sequence has been retrieved from GenBank accession number AF158748. Nucleotide numbering is according to Horikawa *et al.* (2000).

4.4.6 Haplotype frequencies at the calpain 10 locus

Haplotype and haplotype combinations were generated in the following manner. In the case of the CAPN10 gene the order in which the alleles were presented within the haplotype was according to the following template: UCSNP-43, UCSNP-56 and UCSNP-63. An individual homozygous for the 1 allele at UCSNP-43, heterozygous for UCSNP-56 and homozygous for allele 1 at UCSNP-63 would have a 111/121 haplotype. The haplotype combinations were generated according to the list presented in Appendix A. The generation of haplotypes for ambiguous loci is presented in Section 5.4.6.

4.4.7 Detection of single nucleotide polymorphisms in the adiponectin gene

Various genetic alterations within the APM1 gene have been associated with increased susceptibility to T2D within the European population (Schwarz *et al.*, 2004), namely C-11377G, T45G and G-11391A, and were screened in the population under investigation via real-time PCR and melting curve analysis. The positions of C-11377G and G-11391A are indicated within the promoter region of the APM1 gene (GenBank accession number AF304467) whereas the location of T45G is specified within exon 2 of the APM1 gene (GenBank accession number AC072018). All sequence presented in the following sections is indicated in the forward direction.

4.4.8 C-11377G in the adiponectin gene[1]

The C-11377G alteration is a cytosine (allele 1) to a guanine (allele 2) alteration at nucleotide position 8152. The positions of the primer set, ACR1F and ACR1R, and hybridisation probes, SNP1anchor and SNP1sensor, are presented in Table 4.12. Hybridisation probes were designed to recognise the wild type allele. The primer and probe annealing sites utilised in the detection of the G-11391A alteration are also presented in Table 4.12 however this SNP is discussed in greater detail in Section 4.4.10.

Table 4.12: Partial sequence of the adiponectin gene from nucleotide 7990 to 8289

Nucleotide number	DNA Sequence: C-11377G and G-11391A
7990	tgtgtggact gtggagatga tatctggggg gcaggcagac acttgccctg cctctgtctg
8050	agaaaattct gttttggatg tcttgttgaa gttggtgctg gcatcctaag cccttgctgg
8110	ggtcgtaatt taattcatca gaatgtgtgg cttgcaagaa cgggctcaga tcctgcgctt
8170	caaaaacaaa acatgagcgt gccaagaaag tccaaggtgt tgaatgttgc cacttcaagc
8230	ctaaactttc taggaacacc taagtgggtg gcagcttcca gttctccagg ctgcttctag

The forward primer sequence is indicated by the double underlined text (vvv) whereas the position of the reverse primer sequence is indicated by the underlined text (www). The anchor probe sequence specific for C-11377G is indicated by the red text (xxx) whereas the anchor probe specific for G-11391A is indicated by the text highlighted in green (yyy). The sensor probe specific for C-11377G is indicated by the blue text (yyy) whereas the sensor probe specific for G-11391A is indicated by the text highlighted in yellow (zzz). The position of C-11377G is indicated by a block (□) and that of G-11391A is indicated by a circle (○). This sequence has been retrieved from GenBank accession number AF304467. Nucleotide numbering according to Das *et al.* (2000).

The two SNPs were detected utilising different hybridisation probes designed to be specific for each of the alterations. Due to the close proximity of these two alterations the primers ACR1F and ACR1R presented in Table 4.5, were utilised in the PCR amplification of this region for investigation of both alterations.

4.4.9 T45G in the adiponectin gene

The polymorphism termed T45G is a thymine (allele 1) to a guanine (allele 2) alteration at nucleotide position 5629. The primer set ACR3F and ACR3R presented in Table 4.5 was utilised in the amplification of this alteration whereas detection was achieved via the probe set SNP3newss and SNP3newar depicted in Table 4.7. This probe set specifically recognised the variant sequence of this SNP. The sequence has been altered from a thymine to a guanine at position 5629 within Table 4.13. The probe and primer recognition sites are depicted in Table 4.13.

[1] Half of the controls and cohorts described in Section 4.1.1 were genotyped by U. Buro at this locus utilising a protocol optimised by the author.

Table 4.13: Partial sequence of the adiponectin gene from nucleotide 5454 to 5993

Nucleotide number	DNA Sequence: T45G
5454	tacttagaaa gcagctccta gaagtagact ctgctgagat ggacggagtc ctttgtaggt
5514	cccaactggg tgtgtgtgtg gggtctgtct ctccatggct gacagtgcac atgtggattc
5574	cagggctcag gatgctgttg ctgggagctg ttctactgct attagctctg cccgggcatg
5634	accaggaaac cacgactcaa gggcccggag tcctgcttcc cctgcccaag ggggcctgca
5694	caggttggat ggcgggcatc ccagggcatc cgggccataa tggggcccca ggccgtgatg
5754	gcagagatgg cacccctggt gagaagggtg agaaaggaga tccaggtaag aatgtttctg
5814	gcctctttca tcacagacct cctacactga tataaactat atgaaggcat tcattattaa
5874	ctaaggccta gacacaggga gaaagcaaag ctttttttatg ttaaccataa gcaacctgaa
5934	gtgatttggg gttggtcttc caaggatgag tgtagatggt gcctctataa ccaagacttt

The forward primer sequence is indicated by the double underlined text (vvv) whereas the position of the reverse primer sequence is indicated by the underlined text (www). The anchor probe sequence is indicated by the red text (xxx). The sensor probe specific for T45G is indicated by the blue text (yyy). The position of T45G is indicated by a circle (○). This sequence has been retrieved from GenBank accession number AC072018. Nucleotide numbering according to Muzny *et al.* (2000).

4.4.10 G-11391A in the adiponectin gene[1]

The G-11391A alteration is a guanine (allele 1) to adenine (allele 2) alteration at nucleotide position 8166. As stated in Section 4.3.2.1 this SNP and C-11377G are near to each other and therefore the same primer set was utilised to amplify this region. A different set of probes namely AnchorAPM10-1 and SensorGsnp10 listed in Table 4.7, were utilised to detect this alteration. The recognition sites of these probes as well as the primer set are presented in Table 4.12.

4.4.11 Haplotype frequencies at the adiponectin locus

To determine the frequencies of the haplotypes for the genes investigated, within the various populations, it was necessary to elucidate the various haplotype combinations that were present. Haplotypes generated for alterations in the APM1 gene are presented in the following order: C-11377G and G-11391A. Hence an individual homozygous for the 2 allele at C-11377G and homozygous for the 1 allele at G-11391A was assigned the haplotype combination 21/21. Further discussion regarding the utilisation of these specific SNPs is presented in Section 5.4.11.

[1] The black Southern African patients and controls described in Section 4.1.1.1 were genotyped at this locus by U. Buro utilising a protocol optimised by the author.

CHAPTER FIVE

RESULTS AND DISCUSSION

As the populations investigated in this study are already struggling with the added impact of HIV and AIDS (Benatar, 2004) it becomes apparent that health care services cannot afford expensive T2D treatment strategies. This is ironic because disease risk is almost obliterated by the application of the relatively inexpensive and complementary therapies of dietary management and an exercise regimen in a timely fashion (Eriksson and Lindgärde, 1991). The genetic aetiology of this group of disorders was investigated at specific loci in this study in order to describe the possible origins of disease risk. Upon elucidation of these pathways it will be possible to develop pre-diagnostic testing in order to define an individual risk profile to facilitate the implementation of treatment strategies.

The major objective of this research project, however, was to underline the importance of including population specific variability when investigating the role of genetic loci in T2D disease susceptibility. Furthermore, it was undertaken in order to elucidate the genetic factors at certain loci, which increase susceptibility towards T2D in the populations investigated.

5.1 STUDY DESIGN

This report represents the largest investigation into the association of certain SNPs within the CAPN10 and APM1 genes with T2D in the black Southern African and Cuban populations. Individuals were collected according to their diabetic status and were therefore heterogeneous in terms of ethnicity. To overcome possible stratification due to ethnicity a case-control strategy was utilised in order to minimise this effect. Individuals were therefore not sub-classified according to their ethnic origins, but rather grouped according to their diabetic status.

5.1.1 Participant selection

Although there is a close link between T2D and hypertension due to their linked origin to insulin resistance (Reavan, 1988) it was decided to include hypertensive patients as control individuals in this investigation. These individuals were the only available comparable group and were from the same community as the diabetics included in this

investigation. Another reason for their inclusion is that these individuals are monitored for T2D risk and their diabetic status is therefore generally a matter of record. If the controls had been collected through another outpatient clinic or randomly selected from the population the risk of including an undiagnosed diabetic individual would have been higher.

5.1.1.1 Black Southern African cohort

The decision to utilise the black Southern African population was due to the fact that this cohort belongs to macrohaplogroup L. This major haplogroup is the most ancient and therefore also harbours the most variation (Chen *et al.*, 1995). Thus it is an excellent candidate for investigation as it harbours the ancestral frequencies of the various genotypes, haplotypes and haplotype combinations discussed in Section 3.3.1 and 3.3.4. All other populations are derived from this lineage (Cann *et al.*, 1987). Secondly this population is currently undergoing increased risk towards the symptoms of the metabolic syndrome due to urbanisation (Schutte *et al.*, 2003). A recent investigation has determined that the mitochondrial genome has undergone natural selection due to varying climatic conditions and diet as the various migrant populations left Africa (Mishmar *et al.*, 2003). It therefore follows that metabolism is highly attuned to the environment in which an individual is located. Similarly if that environment changes, such an individual is at a disadvantage as its metabolism is no longer optimal in the new environment. This could result in the various lifestyle disorders such as T2D and CAD. This population is therefore at a greater risk of developing these types of disorders. This increased risk and the ancestral nature of this population signifies that it is an excellent candidate for this type of analysis.

5.1.1.2 Cuban cohort

The Cuban population is however more modern in its origins compared to black South Africans. In an investigation by Torroni *et al.* (1995) it was determined that the mitochondrial ancestry of the Cuban population consisted of a heterogeneous mixture with 50% of the mitochondrial haplotypes present, originating from Europeans, 46% originating from Africans and 4% from Native Americans. Whereas the African population harbours the ancestral frequencies of the various genotypes, haplotypes and haplotype combinations under investigation as discussed in Section 5.1.1.1, the current frequencies present in the Cuban population have originated from the initial frequencies within the Native American, African and European populations from which this population has arisen. Therefore if similar patterns of disease risk were defined within these populations it would

lend evidence to the utilisation of global strategies to uncover T2D preventative strategies because ethnic origins would not play a significant role in disease risk. However if this was not the case and the risk factors were different there is a greater likelihood that ethnicity is important in disease risk.

Furthermore this population is not immune to the effect of Westernisation as was determined upon investigation of first generation Cuban immigrants to the United States of America. These individuals were determined to present with greater levels of overweight upon adoption of the Westernised diet than individuals that maintained the ancestral diet (Gordon-Larsen *et al.*, 2003). The prevalence of diabetes within Cuba has been forecast to increase from 5.4% to 8.3% by the year 2025 (King *et al.*, 1998). If the effects of Westernisation as described by Gordon-Larsen *et al.* are taken into account this increase is not surprising. Therefore the Cuban population is a good candidate for genetic elucidation of T2D disease risk.

5.2 ISOLATION OF DNA

The samples utilised were obtained with informed consent from both the patients and controls. A modified protocol of the NucleoSpin® kit was implemented for DNA isolation as discussed in Section 4.2. The average concentration of DNA yielded was ca. 150 $ng.\mu l^{-1}$. Although this technique supplied a relatively lower yield than could be achieved utilising isolation kits which did not utilise spin columns, it was deemed more important to ensure retention of the DNA at the expense of the yield, than to possibly lose the DNA pellet during a washing step. The DNA isolated via this method was clean (average $A_{260}/A_{280} = 1.85$) and of sufficient concentration to be utilised in the numerous downstream molecular screening techniques.

5.3 STATISTICAL ANALYSES

The data generated in this investigation was analysed statistically according to the following methods as described in Section 4.3. Statistical calculations were undertaken utilising the Microsoft™ Excel software. Due to the multiple testing of the cohorts investigated an adjusted significance level had to be defined via the Bonferroni method as presented in Section 4.3.1 (Bland and Altman, 1995).

The genotypes produced were tested for adherence to HW equilibrium (Hardy, 1908) as discussed in Section 4.3.2 whereupon determination of any significant difference in the

frequencies of the genotypes, haplotypes and haplotype combinations between the diabetic and control cohorts investigated was achieved via the use of contingency table analyses and χ^2 testing as outlined in Section 4.3.3 (Samuels *et al.*, 1989). The origins of any significant difference were investigated via odds ratio determination as defined in Section 4.3.4 (Bland and Altman, 2000) whereas the elucidation of any global associations between the specific genotypes and disease risk was achieved via meta-analysis of the Cuban, black Southern African and a reported German population as indicated in Section 4.3.5 (Jackson, 2004).

5.3.1 Calculation of significance level

In defining a threshold value for all statistical analyses, the effect of multiple significance testing needs to be addressed. As discussed in Section 4.3.1 multiple genetic screenings of the Cuban and Southern African cohorts were undertaken. Numerous statistical tests on the same sample increase the probability of a Type I error (Bland and Altman, 1995), which will result in incorrect acceptance of spurious association. In order to overcome this, the significance level was adjusted utilising the Bonferroni method as presented in Section 4.3.1. Upon adjusting for multiple testing however the probability of a Type II error occurring is greatly increased in that a significant association may be overlooked as the significance level is too high (Feise, 2002). Thus in describing a threshold value caution must be applied so as to overcome both of these pitfalls. It was hence decided that a κ value of seven would be applied to the Bonferroni equation at an α value of 0.05, as each cohort underwent analysis of seven alterations. The adjusted significance level of 0.007 was therefore accepted.

5.3.2 Hardy Weinberg equilibrium

Upon determination of the significance level it was possible to define the critical χ^2 value according to a published table described by Samuels *et al.* (1989). The df were determined utilising the equation presented in Equation 4.5 and therefore depended on the size of the contingency tables utilised. Two by three contingency tables as utilised in the analysis of HW equilibrium had 2 df, therefore the critical χ^2 value was defined as 9.21. In a similar manner critical χ^2 values were determined for the larger tables. These relevant values are presented within the footnote of each table as is evident in Table 5.30.

HW equilibrium testing is integral to any case-control analysis investigating genetic variants. This strategy as described in Section 4.3.2 is utilised in order to elucidate

whether the distribution of genotypes within a population are similar to the mathematically derived genotype frequencies as calculated from the allele frequencies (Hardy, 1908). If this is the case certain assumptions can be made regarding the sampling of the individuals within a specific cohort. Firstly the population is significantly large enough to allow for random mating. No genotype is therefore proportionally more prevalent as would be the case in a population undergoing inbreeding, for example, wherein a specific genotype has a skewed proportion due to the increased consanguinity (Hardy, 1908). Furthermore if the population is large enough the effect of random genetic drift is negligible (Degos *et al.*, 1979). Secondly the locus is not undergoing allele reversion, as this would alter allele frequencies, thus altering the calculated genotypic frequencies. Thirdly the population from which the cohorts are collected has not recently undergone migration, causing stratification within a seemingly homogenous population. Allele frequencies may be stable within both original populations but upon migration the allelic levels would be that of the mean of the two populations unless this stratification is corrected (Degos *et al.*, 1979). Finally the population is not experiencing natural selection at the investigated locus. If this had been present it would have resulted in a specific genotype being preferentially increased or decreased between generations ultimately changing allele frequencies and thus preventing HW equilibrium from being reached (Bowcock *et al.*, 1991).

HW equilibrium testing is thus a measure of the amount of variation inherently present in the population, and may affect any association detected. It can be hypothesised that any differences between two cohorts that are both in HW equilibrium, which differ only in disease status, are not due to any of the factors previously discussed and are therefore more likely brought about by association of a specific genotype to disease susceptibility. However, a cohort can only be in HW equilibrium if the amount of variation between the actual genotype frequencies and the calculated frequencies does not exceed a specific threshold value as defined in the first paragraph of Section 5.3.2.

At the outset, it may seem counterintuitive that the population should be in HW equilibrium at these loci, as these variations should undergo natural selection i.e. the changes have a negative effect on the patient's phenotype. However the nature of these alterations, that is their presence at susceptibility loci, results in the disorder generally manifesting after reproductive age, therefore not affecting the fecundity of the individual. For this reason natural selection has very little effect at these loci. The assumptions for testing of the hypothesis therefore remained intact.

Another possible problem in utilising HW equilibrium in this instance is that it requires the individuals in the study to be from the same generation i.e. it requires the existence of discrete generations. This factor was, however, overcome in the population investigated in this study, as one of the criteria for inclusion was that patients were over the age of twenty-five, therefore selecting for a single generation.

To date this is the largest investigation of the allele variation present within the black Southern African and Cuban populations for SNPs involved in T2D susceptibility in comparison to various other populations. HW equilibrium status was calculated for all the cohorts to verify that all groups investigated were representative samples of the greater population under investigation.

5.3.3 Chi square analysis

χ^2 analysis was utilised for comparison of the various control and diabetic cohorts for the calculation of any differences in genotype distribution following HW equilibrium testing. It was further utilised to define any significant differences in the haplotypes and haplotype combinations between the two cohorts. The relevant df and critical χ^2 values were determined as discussed in Section 4.3.3 and Section 5.3.2

5.3.4 Odds ratio determination

Utilising χ^2 analysis it was possible to determine if a significant difference was present in the distributions of the genotypes, haplotypes and haplotype combinations between the cohorts investigated. This was however, merely a measure of the level of difference rather than being indicative of the source of the effect. Therefore it was necessary to utilise a statistical analysis such as OR as discussed in Section 4.3.4 in order to determine the risk of a specific genotype, haplotype or haplotype combination within the cohort (Bland and Altman, 2000).

Woolf's method of calculation was utilised as the populations investigated are unlikely to be genetically homogeneous. This method utilises actual incidence rates as opposed to the frequencies of certain factors within the population. It is, therefore a direct measure of the disease risk within the specific population, which does not require the initial elucidation of differences in frequencies between the different populations investigated for its calculation (Woolf, 1955).

The method described is utilised to calculate the risk OR of the genotype, haplotype or haplotype combination with regard to T2D. It was utilised to compare the evidence for association of a specific factor to disease risk versus the evidence that the factor is associated with decreased risk or protection towards the disease (Bland and Altman, 2000). An OR value of > 1 was taken as an indication that the evidence for association to disease risk was greater, therefore with reference to Equation 4.6 n_1 and n_4 had larger sample sizes than cohorts n_2 and n_3. Alternatively an OR < 1 allows for the assumption that the evidence was stronger for association of decreased risk to a specific factor in which case the opposite is true i.e. n_2 and n_3 are larger than n_1 and n_4. If the cohort under investigation was truly representative of the larger population under investigation the OR was a measure of the risk that a person harbouring a specific factor in the greater population would have of developing the disorder.

The natural logarithm of OR is, however, the central value of a logarithmic distribution from which the 95% CI is generated as indicated in Equation 4.7, and any conclusions made from this former value were always undertaken with reference to the CI. The range of the confidence interval was utilised to infer the likelihood of association between a specific genotype, haplotype or haplotype combination and disease risk:

- 95% CI < 1 : association with a protective effect (McGeer *et al.*, 1996)
- 95% CI < 1 and 95% CI > 1: no association with either an increased or decreased risk (Bland and Altman, 2000)
- 95% CI > 1: association with risk towards the disease (Chrichton, 2001)

These cut off points were utilised as guidelines and any ambiguous intervals are appropriately discussed in the text.

5.3.5 Meta-analysis

In order to determine if any association that was determined at a cohort level translated to a global association, a meta-analysis as discussed in Section 4.3.5 was undertaken at each of the loci investigated. In order to achieve this the black Southern African and Cuban cohorts were analysed with two German cohorts, one of which was screened for the alterations in the CAPN10 gene (Song *et al.*, 2004; Weedon *et al.*, 2003) and one in which the alterations in the APM1 gene were screened (Schwarz *et al.*, 2004). These investigations were chosen, as the individuals were collected in a similar fashion as those utilised in this investigation.

A standard meta-analysis strategy for a case-control study was utilised testing the adherence of the specific genotypes to the model of dominant, recessive or additive effects. The OR generated under the dominant model was regarded as evidence that the association for both the heterozygous and wild type homozygous genotypes to disease risk were similar i.e. disease risk was similarly altered for individuals harbouring these genotypes as compared to individuals homozygous for the mutant allele. Under the recessive model it was possible to determine the effect the wild type homozygote had on disease risk as compared to the other two genotypes i.e. does the wild type allele only alter disease susceptibility in the homozygous state. Finally the additive model was utilised to define the specific effects that each possible genotype had on disease risk if the confounding effects of the other genotypes were removed. Three classes were defined:

- harbouring two wild type alleles versus one (1,1 versus 1,2)
- harbouring one wild type allele versus none (1,2 versus 2,2)
- harbouring two wild type alleles versus none (1,1 versus 2,2).

Since only three populations were investigated it was difficult to determine the level of study heterogeneity via standard graphical methods. It was therefore not possible to determine whether a hypothesis of fixed effects or one of random effects should be utilised. The fixed effects hypothesis allows for the calculation of a measure of whether there is an association of a specific genotype to disease risk in the groups investigated whereas in the case of the random effects hypothesis it is possible to define whether the genotype is always associated with disease risk (Jackson *et al.*, 2004). Calculations for both hypotheses were undertaken at each locus. The Wolfe's method utilised to determine the global OR under fixed effects allows for the generation of $Q_{wolfe,}$ which gives an indication of the heterogeneity across the cohorts, that is ultimately utilised to calculate the OR_{DL}.

Upon calculation of the OR values under these two effects it was possible to determine if the difference was significant via χ^2 testing. If this was the case there was greater evidence for the random effects hypothesis whereas if the values are similar the hypothesis of fixed effects is more appropriate (Jackson *et al.*, 2004). In testing the fixed effects hypothesis, Wolfe's method was utilised due to its relative simplicity as well as the fact that it allows for the comparison of continuous variables, which is useful in investigating complex diseases, which may have environmental effects, which mitigate disease risk. The random effects model was calculated utilising the method described by DerSimonian and Laird, as this is the standard method for meta-analytical analysis under this hypothesis (DerSimonian and Laird, 1986).

5.4 MUTATION DETECTION VIA REAL-TIME POLYMERASE CHAIN REACTION AND MELTING CURVE ANALYSIS

Due to the large cohorts of individuals to be genotyped it was necessary to utilise a fast, simple and time effective method of molecular screening which did not require the use of any dangerous and/or poisonous reagents. Thus it was decided that the alterations within the CAPN10 and APM1 genes would be screened via the utilisation of the LC™ Real Time PCR Machine implementing hybridisation probe technology for genotype detection utilising the phenomenon of FRET.

This method of analysis was chosen due to its many advantages. Firstly, this method of mutation detection is extremely rapid as the sample is able to undergo 30-40 cycles of the PCR within 20-30 min with an added 10-20 min for the mutation detection cycle. Secondly, this process is extremely specific due to the process of mutation detection being based on FRET technology. Finally, it is non-radioactive and non-carcinogenic and negates the necessity for the use of harmful chemicals i.e. ethidium bromide or acrylamide during electrophoresis of the PCR product.

The LC™- FastStart™ Plus DNA Master Hybridisation Probes kit was utilised in the screening of the various genetic alterations within the CAPN10 and APM1 genes to prevent the non-specific amplification, which may interfere with signal detection. It achieves this due to the "hot-start" nature of the polymerase. The enzyme has been treated with a monoclonal antibody which inhibits the FastStart™ polymerase activity and which is only released upon exposure to a temperature of 95°C for ca. 10 min (Kellog *et al.*, 1994). Hence the cycling conditions employed were altered accordingly. Furthermore this kit was utilised as it negated the need for complex optimisation of the reaction parameters as the concentration of magnesium chloride or [$MgCl_2$] had been optimised for allele detection type of analyses.

Each set of reactions contained a negative control in which the aliquot of genomic DNA was replaced by the same volume of double distilled water. This was undertaken, as the LC™ instrument is sensitive to contamination. Furthermore this is good lab practice, in order to ensure the validity of results obtained. As discussed in Section 4.4 there was no need for the utilisation of a uracil-DNA glycosylase as no carry over contamination occurred. This was mainly due to the fact that real-time PCR preparation was undertaken

in a separate room from the post PCR area. Furthermore reagents and equipment utilised for pre- and post PCR analysis were kept separate, according to good laboratory practise.

Each reaction set also included three positive controls in which the DNA aliquot was of control DNA known to harbour one of the three possible genotypes i.e. homozygote for the 1 allele, heterozygote or homozygote for the 2 allele. The graphs that the aforementioned reactions generated were utilised in the standardisation of the differential graphs produced via mutation detection analysis. Alternatively stated, a patient was only scored as harbouring a specific genotype if the resulting peak was at the same temperature as that of one of the control peaks.

During the course of this investigation it was determined that a half of the reaction volume could be utilised in the screening process. The decreased volume still generated a signal with sufficient intensity to be utilised in the exact determination of the genotype under analysis. This revised reaction volume was therefore instituted due to its cost-effectivity.

Upon comparison of the sizes of the different sample groups it is evident that total sample sizes differ for the various SNPs that have been investigated as can be determined upon comparison of Table 5.1 and 5.8. This is due to the fact that as a large quantity of SNP data was generated it was not always possible due to budgetary constraints to repeat all the analyses that did not work the first time. Although cohorts of the exact same sizes are preferential the results determined by the utilisation of different sized cohorts are still significant as long as the cohorts are representative of the larger population. In this investigation this representivity was determined by adherence to HW equilibrium. Due to the low frequencies of certain genotypes such as the 2,2 homozygote at the C-11377G locus within the black Southern African cohorts, it was decided that all the data points generated should be included.

5.4.1 Detection of single nucleotide polymorphisms within the calpain 10 gene

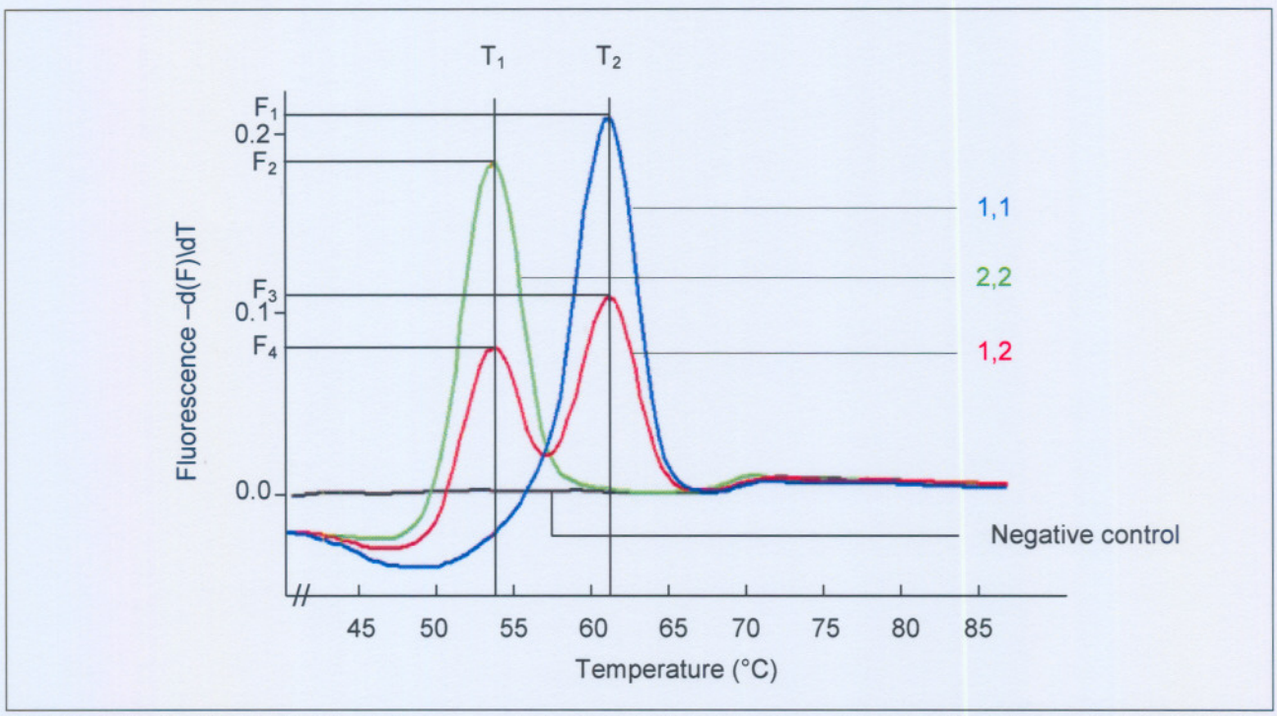
CAPN10 is one of the most investigated loci with regard to T2D. Research into the role of this locus in disease susceptibility has been undertaken in the Mexican American, German (Horikawa *et al.*, 2000), South Indian (Cassell *et al.*, 2002), English (Evans *et al.*, 2001), Polish (Malecki *et al.*, 2002), Samoan (Tsai *et al.*, 2001), Mbuti and Biaka populations (Fullerton *et al.*, 2002) in the past. However this is the first reported investigation into the association of specific SNPs in this gene with disease susceptibility in the black South African and Cuban populations.

Within the following sections the molecular screening of UCSNP-43, -44, -56 and -63 will be discussed. Detection of these alterations was achieved via the methods discussed in Section 4.4. Specifics regarding the thermal cycling program as well as the scoring process are discussed under each heading.

5.4.2 UCSNP-43 within the calpain 10 gene

The amplification of intron three of the CAPN10 gene was undertaken according to the cycling conditions stated in Section 4.4. This reaction was undertaken at an optimised T_a of 58°C for UCSNP-43. It was subjected to 40 cycles of the thermal parameters previously discussed in Table 4.6. Melting peaks were determined via the method discussed previously in Section 4.4 and are presented in Figure 5.1. Due to the variation of the melting temperatures at which the peaks were produced, the positive controls were utilised for calibration purposes.

Figure 5.1: Diagrammatic representation of the differential graph of probe fluorescence versus temperature for UCSNP-43 within the calpain 10 gene



UCSNP-43: 1 = G allele, 2 = A allele; 1,1 = homozygote for the 1 allele indicated by the blue line; 1,2 = heterozygote indicated by the red line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by black line wherein DNA was replaced by deionised water; T_1 = temperature at which the 2,2 homozygote peaks (T_1 = 54.15°C, standard deviation = 0.34 °C); T_2 = temperature at which the 1,1 homozygote peaks (T_2 = 61.57°C, standard deviation = 0.71°C); F_1 = level of differential fluorescence of the peak of the 1,1 homozygote; F_2 = level of differential fluorescence of the peak of the 2,2 homozygote; F_3 = level of differential fluorescence of the right hand peak of the 1,2 heterozygote; F_4 = level of differential fluorescence of the left hand peak of the 1,2 heterozygote; As the y axis represents the negative differential of fluorescence with regard to time there are no units presented.

The probe for UCSNP-43 specifically recognises the G allele. The peak with the highest melting temperature (T_2 = 61.57°C, standard deviation = 0.71°C) as depicted in Figure 5.1 (blue curve), was designated the 1,1 homozygote whereas the peak at the lower temperature (T_1 = 54.15°C, standard deviation = 0.34°C) was classified as the 2,2 homozygote (green curve in Figure 5.1). If there were two peaks present as represented by the red curve in Figure 5.1, both had to correlate with the melting temperatures of the 1,1 and 2,2 genotypes for the patient to be scored as a heterozygote (heterozygote T_1 = 54.40°C, standard deviation = 0.52°C; heterozygote T_2 = 61.96°C, standard deviation = 0.55°C). Furthermore the amplitude of the two peaks had to be approximately half that of the peak of the homozygote. In Figure 5.1 for example the level of fluorescence of the red curve at T_1 (F_4) is lower than that at T_2 (F_3). However the green curve representing the 2,2 homozygote itself has a lower amplitude (F_2) than the blue curve representing the 1,1 homozygote (F_1). The difference between F_2 and F_1 is similar to that of F_4 and F_3 and thus it is possible to define the red curve as a heterozygote.

5.4.2.1 UCSNP-43 within the black Southern African diabetic cohort

Via the utilisation of the methods discussed in Section 4.4 it was possible to determine the genotypic frequencies of UCSNP-43 within the CAPN10 gene. As depicted in Table 5.1, UCSNP-43 is in fact in HW equilibrium (χ^2 = 2.87) within the South African diabetic cohort under investigation. The population is thus unaffected by the various factors discussed in Section 5.3.2.

Table 5.1: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African patient cohort for UCSNP-43

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	183	48	7	238
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.77	0.20	0.03	1.00
Expected numbers (E)	180.04	53.92	4.04	238.00
$\chi^2 = (O-E)^2/E$	0.05	0.65	2.17	2.87

UCSNP-43: 1 = G allele, 2 = A allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.2.2 UCSNP-43 within the black Southern African control cohort

It was necessary to determine whether the various control cohorts were in HW equilibrium at the various loci previously discussed. By this calculation it was determined that the Southern African control cohort was in fact in HW equilibrium at UCSNP-43 ($\chi^2 = 1.52$) as depicted in Table 5.2. The absence of the effects delineated in Section 5.3.2 allows for the statistical comparison of the black Southern African control and affected groups at this locus.

Table 5.2: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African control cohort for UCSNP-43

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	170	58	2	230
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.74	0.25	0.01	1.00
Expected numbers (E)	172.18	53.64	4.18	230.00
$\chi^2 = (O-E)^2/E$	0.03	0.35	1.14	1.52

UCSNP-43: 1 = G allele, 2 = A allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.2.3 Comparison of UCSNP-43 between both black Southern African cohorts

UCSNP-43 was determined to present with a significant difference (p value < 0.001) between the patient and control cohorts investigated as indicated in Table 5.3. It was however not possible to discern the origins of the effect, as upon calculation of the odds ratio, detection of an increased risk for any of the genotypes investigated at the 95% CI included both numbers below and above one, thus indicating neither a protective factor, nor a risk factor as discussed in Section 5.3.4. The highest odds ratio value is that of the homozygote of the variant allele as indicated by the shaded cell in Table 5.3 but possibly due to the relative rarity of this allele, the 95% CI is too broad. This in turn increases the probability of a type 1 error above acceptable limits.

Table 5.3: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-43 locus between the black Southern African patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.77 (n=183)	0.20 (n=48)	0.03 (n=7)	<0.001
Control	0.74 (n=170)	0.25 (n=58)	0.01 (n=2)	
OR (95% CI)	1.17 (0.77-1.78)	0.74 (0.48-1.15)	3.45 (0.71-16.8)	

UCSNP-43: 1 = G allele, 2 = A allele; OR = odds ratio; 95% CI = 95% confidence interval; n = sample size; shaded block indicates the evidence for the increased risk associated with the 2,2 genotype.

5.4.2.4 UCSNP-43 within the Cuban diabetic cohort

Investigation of UCSNP-43 in the patient cohort of the Cuban population resulted in the determination of the values as depicted in Table 5.4. The cohort is in HW equilibrium ($\chi^2 = 0.51$) with respect to this specific polymorphism and therefore the population is not affected by the various factors discussed in Section 5.3.2. The absence of migration is heartening as it means the cohort collected has not recently had an addition or subtraction of alleles due to this phenomenon.

Table 5.4: Chi-square test of goodness-of-fit to the HW proportions of the Cuban patient cohort for UCSNP-43

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	200	104	17	321
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.62	0.32	0.05	1.00
Expected numbers (E)	197.83	108.34	14.83	321.00
$\chi^2 = (O-E)^2/E$	0.02	0.17	0.32	0.51

UCSNP-43: 1 = G allele, 2 = A allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.2.5 UCSNP-43 within the Cuban control cohort

The UCSNP-43 locus within the control group of the Cuban population, as presented in Table 5.5, was in HW equilibrium ($\chi^2 = 0.91$). Therefore any difference in genotype frequencies determined upon comparison to the patient cohort could be attributed to association with T2D. At this locus, the numerous modifying effects discussed in Section 5.3.2 are negligible within this cohort.

Table 5.5: Chi-square test of goodness-of-fit to the HW proportions of the Cuban control cohort for UCSNP-43

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	147	96	11	254
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.58	0.38	0.04	1.00
Expected numbers (E)	149.70	90.60	13.70	254.00
$\chi^2 = (O-E)^2/E$	0.05	0.32	0.53	0.91

UCSNP-43: 1 = G allele, 2 = A allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.2.6 Comparison of UCSNP-43 between both Cuban cohorts

As this locus was in HW equilibrium in both the Cuban patient and control cohorts it was possible to compare each to the other utilising χ^2 testing, the results of which are depicted in Table 5.6. The absence of association at UCSNP-43 is important, as this locus may therefore not be an indicator of disease risk. This may be important in the future profiling of risk loci in T2D susceptibility in this population.

Table 5.6: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-43 locus between the Cuban patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.62 (n=200)	0.32 (n=104)	0.05 (n=17)	NS
Control	0.58 (n=147)	0.38 (n=96)	0.04 (n=11)	
OR (95% CI)	1.20 (0.85-1.68)	0.78 (0.55-1.11)	1.23 (0.56-2.68)	

UCSNP-43: 1 = G allele, 2 = A allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; n = sample size.

5.4.2.7 Meta-analysis of UCSNP-43

The variant allele at the UCSNP-43 locus is more prevalent in the various Caucasoid (frequency ca. 0.3), Asian and Amerindian (frequency ca. 0.2) populations than the Southern African cohort which has a frequency = 0.13 (Horikawa *et al.*, 2000; Baier *et al.*, 2000; Evans *et al.*, 2001; Malecki *et al.*, 2002; Cassell *et al.*, 2002). As stated previously in Section 5.4.2.3, the homozygotes for the variant allele presented with a high odds ratio indicating a possible association with a risk factor, but it was not significant due to the wide 95% CI. This may be important since the wild type allele is generally associated with disease risk (Horikawa *et al.*, 2000). The confidence interval could be decreased if a larger sample was investigated however the variant allele is present at such a low frequency in the African cohort presented here, that it may ultimately not be a major factor to diabetic susceptibility in the black Southern African population. Alternatively this allele is present at a similar frequency to the non-African populations as presented in Table 3.2 within the Cuban cohorts investigated (frequency = 0.21). The absence of association to this allele within this cohort however indicates that it may not be a significant risk factor in T2D susceptibility in the Cuban population.

According to the Out Of Africa hypothesis the various non-African populations arose via specific migratory events from the African continent (Cann *et al.*, 1987). It is possible that the variant allele may have undergone purifying selection in the African population thus

resulting in its low frequency. Upon the migration of individuals out of Africa this selection force may have diminished as populations were exposed to environments to which this variant allele was better suited, thus allowing for its enrichment in the non-African populations.

Upon meta-analysis, of this locus within the German, black Southern African and Cuban cohorts as presented in Table 5.7, it was determined that a specific genotype was associated with increased risk only under the additive model as indicated by the shaded block in the aforementioned table. The 1,1 genotype presented with a slight increase in disease risk over the 1,2 genotype which is most likely due to the association of the 1,1 genotype to disease risk in the German population (Horikawa *et al.*, 2000). However the evidence of association is masked upon inclusion of the Southern African and Cuban data.

Table 5.7: Meta-analysis of genotypes at the UCSNP-43 locus for the black Southern African, Cuban and German cohorts

Model	Genotype Comparison	Cohort size (cases/controls)	OR (95% CI)		p-value
			Fixed effects	*Random effects*	
Recessive	1,1 vs 1,2 and 2,2	791/572	1.24 (0.98-1.56)	1.24 (0.98-1.56)	NS
Dominant	1,1 and 1,2 vs 2,2	791/572	0.94 (0.55-1.60)	0.87 (0.41-1.86)	NS
Additive	1,1 vs 1,2	749/549	1.29 (1.01-1.64)	1.29 (1.01-1.64)	NS
	1,2 vs 2,2	288/217	0.80 (0.46-1.39)	0.74 (0.34-1.61)	NS
	1,1 vs 2,2	545/378	1.03 (0.60-1.77)	0.95 (0.42-2.12)	NS

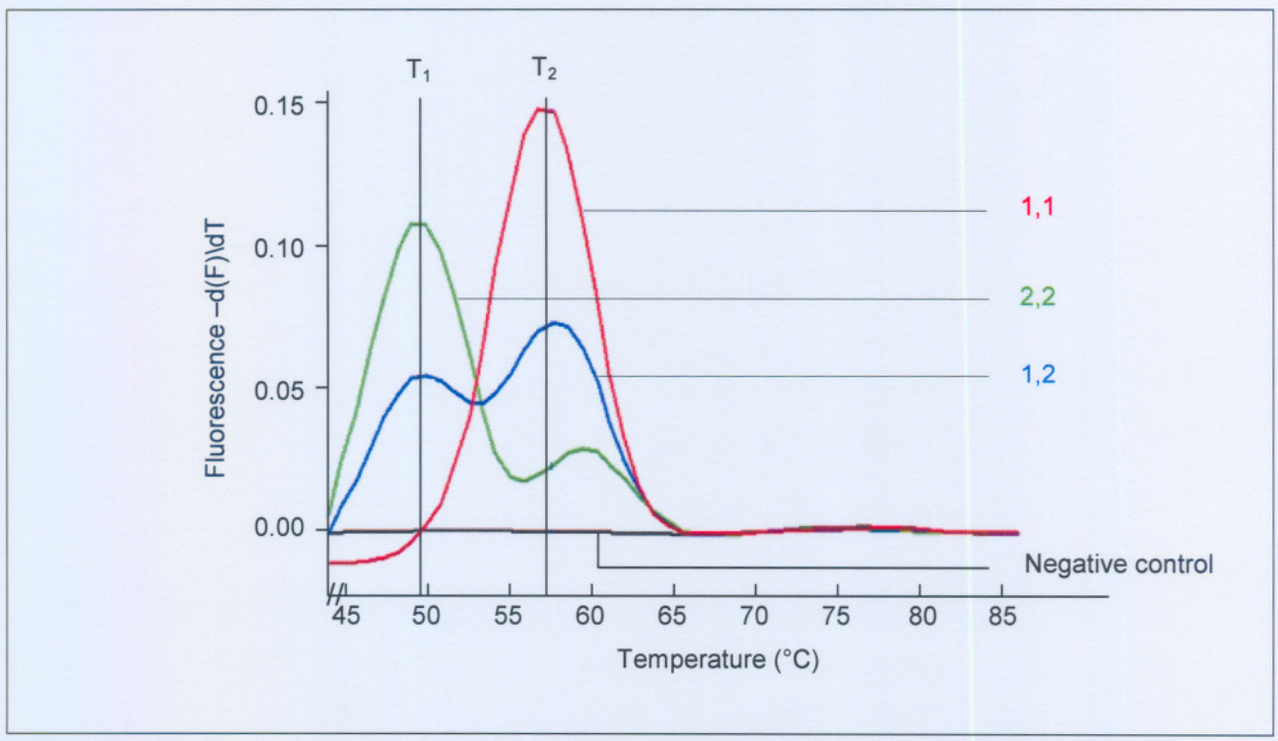
UCSNP-43: 1 = G allele, 2 = A allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; vs = versus; shaded block indicates the evidence for the increased risk associated with the 1,1 genotype versus the 1,2 genotype under the additive model.

Both the random effects and fixed effects models produced identical results thus indicating that the intra cohort variance was greater than the inter cohort variance which signifies a similarity in genotype distribution. However upon analysis of the separate populations as discussed in Section 5.4.2.3 and 5.4.2.6 it is apparent that there are differences in disease susceptibility i.e. the 1,1 genotype is not associated with disease risk in the Cubans and Southern African cohorts investigated but it is in the German cohort. By determining the global effect of a genotype across various ethnic groups, population specific effects may not be detected. Horikawa *et al.* reported the 1,1 genotype to be associated with increased risk for T2D in the European population. However, this association was not maintained upon calculation of its global effect via meta-analysis in this study. This is discussed further in Section 6.4.

5.4.3 UCSNP-44 within the calpain 10 gene

Amplification of intron three which harbours UCSNP-44 was achieved utilising a T_a of 57°C according to the cycling conditions presented in Table 4.6. The sensor probe employed in the detection of alleles at the UCSNP-44 locus specifically recognises the T allele. Thus the melting curves produced for this alteration are depicted in Figure 5.2. The 1,1 homozygote (red curve in Figure 5.2) was assigned to the melting peak at the higher temperature (T_2 = 57.23°C, standard deviation = 1.32°C) whereas the lower peak (T_1 = 49.42°C, standard deviation = 1.27°C) was scored as the 2,2 homozygote (green curve in Figure 5.2). The heterozygote was only scored if there were two peaks corresponding to both the temperatures of the melting peaks generated for the homozygotes, similar to the blue curves in Figure 5.2 (heterozygote T_1 = 49.57°C, standard deviation = 1.24°C; heterozygote T_2 = 57.50°C, standard deviation = 1.30°C). Similar to the discussion in Section 5.4.2 the amplitude of the two peaks generated for a heterozygote had to be ca. half that of the respective peaks generated for the homozygotes.

Figure 5.2: Diagrammatic representation of the differential graph of probe fluorescence versus temperature for UCSNP-44 within the calpain 10 gene



UCSNP-44: 1 = T allele; 2 = C allele; 1,1 = homozygote for the 1 allele indicated by the red line; 1,2 = heterozygote indicated by the blue line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by black line wherein DNA was replaced by deionised water; T_1 = temperature at which the 2,2 homozygote peaks (T_1 = 49.42°C, standard deviation = 1.27°C); T_2 = temperature at which the 1,1 homozygote peaks (T_2 = 57.23°C, standard deviation = 1.32 °C).

5.4.3.1 UCSNP-44 within the black Southern African diabetic cohort

The various assumptions discussed in Section 5.3.2 may be applied to this population as it was calculated that this cohort was in HW equilibrium ($\chi^2 = 0.36$) at this specific locus, as presented in Table 5.8. There was only a single individual harbouring the 2,2 genotype at this SNP. It is therefore plausible that this genotype is present at a very low frequency within the larger population. Although it may be possible to determine an association between this genotype and T2D upon comparison to the control population it may not be relevant as a population level risk factor as too few individuals would harbour this genotype. Furthermore this locus has been determined to have low discerning power within the European and Mexican American populations (Horikawa *et al.*, 2000). A similar situation may be present in this cohort. However, if a significant difference in genotype frequencies between this cohort and the unaffected individuals were determined, it may be due to association between the disorder and this allele.

Table 5.8: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African patient cohort for UCSNP-44

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	211	21	1	233
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.91	0.09	0.00	1.00
Expected numbers (E)	210.57	21.86	0.57	233.00
$\chi^2 = (O-E)^2/E$	0.00	0.03	0.33	0.36

UCSNP-44: 1 = T allele; 2 = C allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.3.2 UCSNP-44 within the black Southern African control cohort

Analysis of the genotype frequencies at UCSNP-44 resulted in the determination that the cohort under investigation was also in HW equilibrium ($\chi^2 = 0.80$) at this locus as presented in Table 5.9. The implications of this are discussed in Section 5.3.2. As previously discussed the variant allele is relatively rare within both the patient and control groups investigated therefore the differentiation power is not very high. This is part of the reason why this locus has been excluded from the analyses of the haplotypes and haplotype combinations.

Furthermore in the original analysis of Mexican Americans (Horikawa *et al.*, 2000) the SNP was only differentiable in association with the 111 haplotype i.e. producing the 1111 and 1211 haplotypes. The three other haplotypes defined within this population were therefore 1121, 2121 and 1122 thus inclusion of this SNP does not add great value to haplotype analysis. It is possible that this alteration may be associated with T2D even though it is at a low level in the population. It was therefore tested via χ^2 analysis as presented in Table 5.9 to 5.11.

Table 5.9: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African control cohort for UCSNP-44

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	199	28	2	229
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.87	0.12	0.01	1.00
Expected numbers (E)	198.12	29.76	1.12	229.00
$\chi^2 = (O-E)^2/E$	0.00	0.10	0.70	0.80

UCSNP-44: 1 = T allele; 2 = C allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.3.3 Comparison of UCSNP-44 between both black Southern African cohorts

There was no significant deviation in the genotype distribution within the Southern African cohorts investigated as indicated in Table 5.10. It is therefore unlikely that there is an association between this SNP and disease risk within the cohort investigated. The low frequency of the variant allele as well as the absence of any significant disease risk indicates that this locus may not be a significant risk factor in the Southern African population.

Table 5.10: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-44 locus between the black Southern African patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.91 (n=211)	0.09 (n=21)	0.00 (n=1)	NS
Control	0.87 (n=199)	0.12 (n=28)	0.01 (n=2)	
OR (95% CI)	1.44 (0.80-2.59)	0.71 (0.39-1.29)	0.48 (0.04-5.43)	

UCSNP-44: 1 = T allele; 2 = C allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; n = sample size.

5.4.3.4 UCSNP-44 within the Cuban diabetic cohort

Elucidation of the genotypes present at UCSNP-44 within the Cuban patient cohort was undertaken and the results are presented in Table 5.11. The population from which the

cohort was sampled is sufficiently large to allow for random mating as the group investigated was in HW equilibrium (χ^2 = 0.90). The numerous other modifying factors as discussed in Section 5.3.2 therefore also do not have an effect.

Table 5.11: Chi-square test of goodness-of-fit to the HW proportions of the Cuban patient cohort for UCSNP-44

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	253	62	6	321
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.79	0.19	0.02	1.00
Expected numbers (E)	251.26	65.47	4.26	321.00
$\chi^2 = (O-E)^2/E$	0.01	0.18	0.71	0.90

UCSNP-44: 1 = T allele; 2 = C allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.3.5 UCSNP-44 within the Cuban control cohort

The UCSNP-44 locus within the control cohort of Cuban origin is in HW equilibrium as determined via χ^2 testing (χ^2 = 0.86) as depicted in Table 5.12. Determination of HW equilibrium within this group further allows that the effects discussed in Section 5.3.2 may be overlooked, if any variation in genotype frequencies is present upon comparison to the patient cohort.

Table 5.12: Chi-square test of goodness-of-fit to the HW proportions of the Cuban control cohort for UCSNP-44

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	196	56	2	254
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.77	0.22	0.01	1.00
Expected numbers (E)	197.54	52.91	3.54	254.00
$\chi^2 = (O-E)^2/E$	0.01	0.18	0.67	0.86

UCSNP-44: 1 = T allele; 2 = C allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.3.6 Comparison of UCSNP-44 between both Cuban cohorts

The absence of any significant alteration in genotype frequency at this locus between the control and diabetic cohorts investigated as presented in Table 5.13, is indicative of a lack of association with disease risk. Although the homozygote for the variant allele presented with a high OR value, the width of the 95% CI is too broad to allow for the generation of any meaningful conclusions regarding this value. Therefore UCSNP-44 is unlikely to be an

important risk factor in the Cuban population if this cohort is truly representative of the greater population. The fact that it is in HW equilibrium however alludes to this being true.

Table 5.13: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-44 locus between the Cuban patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.79 (n=253)	0.19 (n=62)	0.02 (n=6)	NS
Control	0.77 (n=196)	0.22 (n=56)	0.01 (n=2)	
OR (95% CI)	1.10 (0.74-1.63)	0.84 (0.56-1.27)	2.4 (0.48-11.9)	

UCSNP-44: 1 = T allele; 2 = C allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant, n = sample size.

5.4.3.7 Meta-analysis of UCSNP-44

Upon meta-analytical comparison as presented in Table 5.14, it was determined that the UCSNP-44 locus was unlikely to be a significant risk factor. The similarities in the values detected for both fixed effects and random effects indicates that the inter cohort variance is negligible indicating that the cohorts investigated were statistically homogeneous. As no association was present in the individual cohorts this is not unexpected.

Table 5.14: Meta-analysis of genotypes at the UCSNP-44 locus for the black Southern African, Cuban and German cohorts

Model	Genotype Comparison	Cohort size (cases/controls)	OR (95% CI)		p-value
			Fixed effects	*Random effects*	
Recessive	1,1 vs 1,2 and 2,2	862/556	1.08 (0.81-1.44)	1.08 (0.78-1.49)	NS
Dominant	1,1 and 1,2 vs 2,2	862/556	0.69 (0.22-2.13)	0.69 (0.22-2.13)	NS
Additive	1,1 vs 1,2	849/551	1.11 (0.82-1.49)	1.11 (0.81-1.52)	NS
	1,2 vs 2,2	170/103	0.64 (0.20-2.05)	0.64 (0.20-2.05)	NS
	1,1 vs 2,2	705/458	0.69 (0.22-2.15)	0.69 (0.22-2.15)	NS

UCSNP-44: 1 = T allele; 2 = C allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; vs = versus.

The above results are in apparent disagreement with a meta-analysis undertaken by Weedon *et al.* (2003) which investigated UCSNP-44 in the British, German, Japanese, Czechoslovakian and Mexican populations wherein an OR of 1.17 (95% CI 1.07-1.34) was generated indicating the SNP has a small but significant role in T2D susceptibility. However, in the analysis presented by Weedon *et al.* only non-African populations were included. It is therefore possible that inclusion of the Southern Africans and Cubans may have masked this effect in the meta-analysis presented. If this is the case it increases the evidence that susceptibility loci should be investigated at a population level.

5.4.4 UCSNP-56 within the calpain 10 gene

This investigation is part of a larger research program into the elucidation of T2D disease susceptibility. As large groups of SNPs need to be generated it is necessary to implement the highest throughput strategies in order to generate this information. Since an amplified fragment length polymorphism (AFLP) is employed in the detection of UCSNP-19 it would be more time effective to genotype a SNP that is in perfect linkage disequilibrium with this alteration via real-time PCR. The alteration termed UCSNP-56 has been determined to be in perfect linkage disequilibrium with UCSNP-19 in the Mexican American population (Horikawa *et al.*, 2000) and hence this locus was genotyped in this study in accordance with global practise.

This is possibly not the most effective strategy within the African population as presented in an investigation by Fullerton *et al.* in 2002. In this investigation it was elucidated that although linkage between these two SNPs was high in the Chinese, Yakut, Nasioi, Maya and Surui, there was substantially less linkage between these loci within the African cohorts i.e. the Biaka and Mbuti pygmy populations investigated as outlined in Table 5.15.

The Mbuti and Biaka pygmies were the only African populations investigated by Fullerton *et al.* (2002). It is possible that these populations are not representative of all African populations. Evidence for this argument was presented by Watson *et al.* (1996) who determined that these two populations present with approximately 10 times the mitochondrial sequence variation as compared to various other non-pygmy African populations. The Mbuti and Biaka as well as the !Kung populations presented with the most ancestral lineages within their investigation. The lack of linkage disequilibrium may thus be a population specific effect which is not present in other African populations. The Biaka individuals screened in the investigation by Fullerton *et al.* were also previously determined to have undergone high levels of admixture with non-Pygmy Africans (Bowcock *et al.*, 1991). This is a possible explanation for the increased levels of linkage disequilibrium in this population as indicated in Table 5.15, if there are higher levels of linkage disequilibrium within the non-Pygmy African population. If the fact that only 70 Mbuti and 39 Biaka individuals were included is also considered, it is possible that the findings by Fullerton *et al.* (2002) are not representative of all African populations. Future investigations of representative African populations are however necessary to elucidate the exact relationship between the UCSNP-56 and UCSNP-19 loci.

Furthermore, the various non-African population groups within this research program have had the UCSNP-56 locus screened due to the greater logistic feasibility of utilising this strategy as discussed in the previous paragraph. Therefore in order to effectively compare the African cohorts to the non-African cohorts within the larger research program, this strategy was also implemented in the African cohorts. It was deemed that re-genotyping all the individuals for UCSNP-19 would not be cost or time effective.

Haplotypes were thus generated utilising the following order of alterations: UCSNP-43, UCSNP-56 and UCSNP-63. This implied that upon comparison to the reported frequencies of haplotypes and haplotype combinations the haplotype structure was altered accordingly with UCSNP-19 translated to UCSNP-56. Thus the "at risk" haplotype combinations determined in the Mexican American and German populations, namely the 112/121 haplotype combination (Horikawa *et al.*, 2000) and the 121/121 haplotype combination in the Polish population (Malecki *et al.*, 2002) which harbour UCSNP-19 at the centre position, will be represented by 122/111 and 111/111 respectively in order to allow for comparison to the black South African and Cuban cohorts. To prevent confusion both the original and inferred versions are presented in all tables with reference to haplotypes and haplotype combinations as exemplified in Table 5.30.

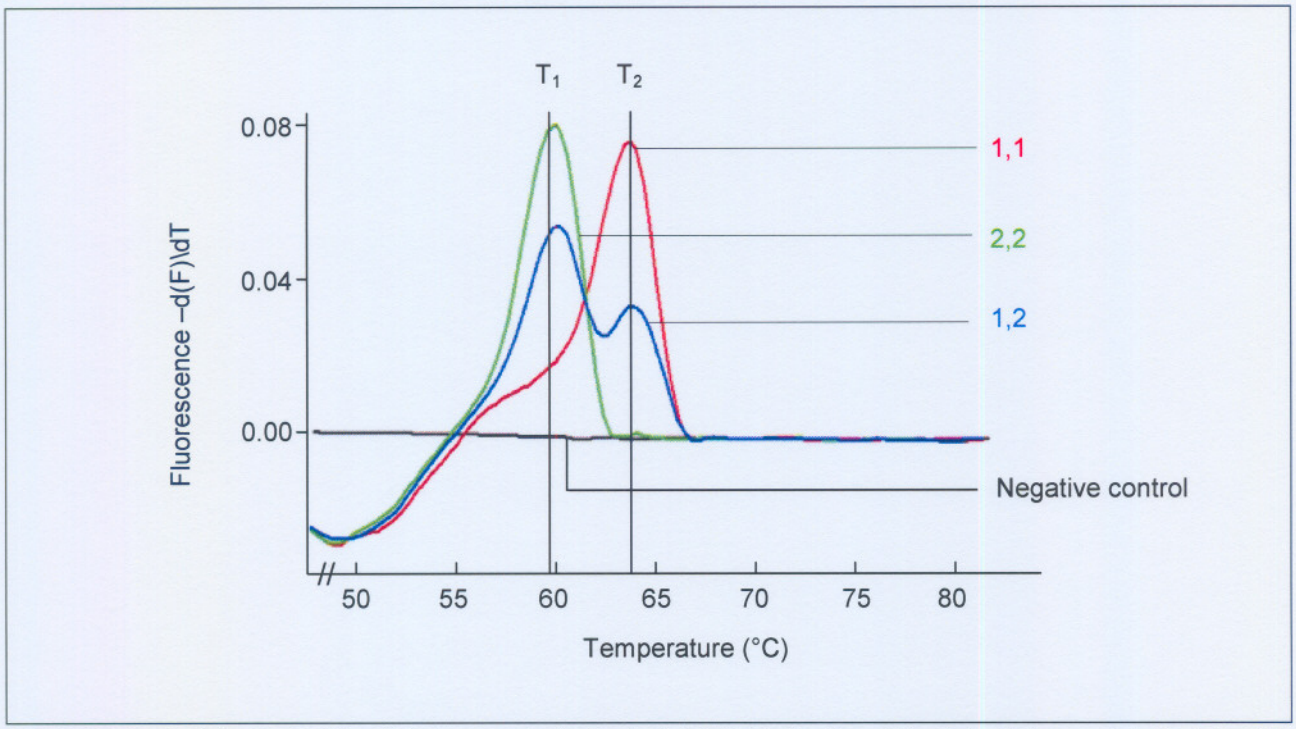
Table 5.15: Linkage disequilibrium between UCSNP-56 and UCSNP-19 in different populations

Ethnic group	Population	Estimation of LD	Ethnic group	Population	Estimation of LD
Native American	Maya	1.000	Asian	Japanese	0.471
	Surui	1.000	European	Druze	0.891
Oceanic	Nasioi	1.000		Danes	0.594
Asian	Chinese	0.960	African	Biaka	0.465
	Nakut	0.846		Mbuti	0.188

LD = linkage disequilibrium. Adapted from Fullerton *et al.* (2002).

Amplification of the region harbouring UCSNP-56 was achieved utilising the reaction conditions as presented in Section 4.4. The reaction was optimised at 40 cycles with a T_a of 58°C for the detection of this specific alteration. Mutation analysis was subsequently achieved via investigation of the melting curve produced by the LightCycler™ software as indicated in Figure 5.3.

Figure 5.3: Diagrammatic representation of the differential graph of probe fluorescence versus temperature for UCSNP-56 within the calpain 10 gene



UCSNP-56: 1 = A allele; 2 = G allele; 1,1 = homozygote for the 1 allele indicated by the red line; 1,2 = heterozygote indicated by the blue line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by black line wherein DNA was replaced by deionised water; T_1 = temperature at which the 2,2 homozygote peaks (T_1 = 59.84°C, standard deviation = 0.36°C); T_2 = temperature at which the 1,1 homozygote peaks (T_2 = 63.91°C, standard deviation = 0.71°C).

Three possible genotype categories could be defined upon inspection of the negative differential graph of fluorescence intensity. As the sensor probe specifically recognises the 1 allele the 1,1 homozygote was represented by a peak at the higher melting temperature (T_2 = 63.91°C, standard deviation = 0.71°C) as represented by the red curve in Figure 5.3 whereas the peak at the lower temperature (T_1 = 59.84°C, standard deviation = 0.36°C) corresponded to the 2,2 homozygote (green curve). A heterozygote was only classified if the graph consisted of two peaks at the representative melting temperatures of the homozygotes as is depicted by the blue curve in Figure 5.3 (heterozygote T_1 = 60.04°C, standard deviation = 0.50°C; heterozygote T_2 = 64.15°C, standard deviation = 0.50°C). Furthermore the amplitude of the two curves had to be approximately half that of the curves generated for the homozygotes as described in Section 5.4.2. All genotype determinations were calibrated according to the peaks of the three positive controls.

5.4.4.1 UCSNP-56 within the black Southern African diabetic cohort

The locus at UCSNP-56 was determined to be in HW equilibrium (χ^2 = 0.95) as depicted in Table 5.16. The cohorts investigated are therefore not currently subjected to any of the effects that were discussed in Section 5.3.2. If the control cohort is similarly in HW

equilibrium it will be possible to compare both groups in order to detect any association between this locus and T2D susceptibility.

Table 5.16: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African patient cohort for UCSNP-56

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	12	94	129	235
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.05	0.40	10.55	1.00
Expected numbers (E)	14.81	88.37	131.8	235.00
$\chi^2 = (O-E)^2/E$	0.53	0.36	0.06	0.95

UCSNP-56: 1 = A allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.4.2 UCSNP-56 within the black Southern African control cohort

UCSNP-56 was determined to be in HW equilibrium ($\chi^2 = 1.64$) within the control group investigated as outlined in Table 5.17. The assumptions associated with this form of equilibrium, as discussed in Section 5.3.2, may be accepted within this cohort at this locus. It will therefore be possible to compare both the control and diabetic cohorts for association between this SNP and T2D.

Table 5.17: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African control cohort for UCSNP-56

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	40	100	89	229
Expected proportions	p^2	$2pq$	q^2	$p^2 \bullet 2pq \bullet q^2$
Frequency	0.17	0.44	0.39	1.00
Expected numbers (E)	35.37	109.26	84.37	229.00
$\chi^2 = (O-E)^2/E$	0.61	0.78	0.25	1.64

UCSNP-56: 1 = A allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.4.3 Comparison of UCSNP-56 between both black Southern African cohorts

The UCSNP-56 alteration was significantly different between the two South African cohorts investigated (p value < 0.0001) as depicted in Table 5.18. Odds ratio determination indicated that the wild type homozygote was associated with a protective effect as discussed by McGeer *et al.* (1996) towards disease risk (OR 0.25, 95% CI 0.12-0.49) as indicated by the lighter shaded block in Table 5.18. Further evidence for this association is

determined upon OR calculation for the 2,2 homozygote which is significantly associated with an increased risk in disease susceptibility (OR 1.91, 95% CI 1.32-2.77) as highlighted by the darker shaded block in Table 5.18. This association of the 1,1 homozygote is important as the various loci within CAPN10 have been reported to be in association with disease risk (Horikawa *et al.*, 2000). Thus it is possible that there is a difference in genetic aetiology of T2D with regard to this locus in the black Southern African population.

Table 5.18: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-56 locus between the black Southern African patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.17 (n=40)	0.40 (n=94)	0.55 (n=129)	<0.0001
Control	0.05 (n=12)	0.44 (n=100)	0.39 (n=89)	
OR (95% CI)	0.25 (0.12-0.49)	0.86 (0.59-1.24)		

UCSNP-56: 1 = A allele; 2 = G allele; OR = odds ratio; 95% CI = 95% confidence interval; n = sample size; the lighter shaded block indicates the evidence for association of the 1,1 genotype with a protective effect; the darker shaded block indicates the evidence for association of the 2,2 genotype to increased disease risk.

5.4.4.4 UCSNP-56 within the Cuban diabetic cohort

HW equilibrium was detected upon genotype screening of UCSNP-56 (χ^2 = 3.53) as presented in Table 5.19. This implies that within this patient cohort any differences determined between this cohort and a control cohort similarly in HW equilibrium may be ascribed to association with the T2D phenotype. Thus any deviations detected would not be due to the aspects explained in Section 5.3.2.

Table 5.19: Chi-square test of goodness-of-fit to the HW proportions of the Cuban patient cohort for UCSNP-56

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	100	143	78	321
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.31	0.45	0.24	1.00
Expected numbers (E)	91.63	159.74	69.63	321.00
$\chi^2 = (O-E)^2/E$	0.77	1.76	1.01	3.53

UCSNP-56: 1 = A allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.4.5 UCSNP-56 within the Cuban control cohort

The genotype frequencies at the UCSNP-56 locus within the Cuban control cohort are not affected by natural selection, migratory events or allele reversion as discussed in

Section 5.3.2. This was concluded upon determination that this locus is in HW equilibrium within the cohort investigated ($\chi^2 = 0.02$) as listed in Table 5.20.

Table 5.20: Chi-square test of goodness-of-fit to the HW proportions of the Cuban control cohort for UCSNP-56

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	82	123	48	253
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.32	0.49	0.19	1.00
Expected numbers (E)	81.39	124.22	47.39	253.00
$\chi^2 = (O-E)^2/E$	0.00	0.01	0.01	0.02

UCSNP-56: 1 = A allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.4.6 Comparison of UCSNP-56 between both Cuban cohorts

As both Cuban cohorts were in HW equilibrium it was possible to compare the genotype distribution of these two groups. Although significant variation was detected in the Southern African cohorts, the Cuban individuals did not present with significant levels of variation at this locus as depicted in Table 5.21. The strong association detected in the Southern African group and the apparent absence of any significant difference between the Cuban cohorts garners greater evidence for population specific effects at T2D susceptibility loci among different populations.

Table 5.21: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-56 locus between the Cuban patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.31 (n=100)	0.45 (n=143)	0.24 (n=78)	NS
Control	0.32 (n=82)	0.49 (n=123)	0.19 (n=48)	
OR (95% CI)	0.94 (0.66-1.34)	0.84 (0.61-1.18)	1.37 (0.91-2.05)	

UCSNP-56: 1 = A allele; 2 = G allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; n = sample size.

5.4.4.7 Meta-analysis of UCSNP-56

The most significant effect detected within the black South African cohort is that of the association between the UCSNP-56 wild type homozygote and protection against T2D (p value < 0.0001, OR 0.25, 95% CI 0.12-0.49) in Section 5.4.4.3. This is important to note, as this allele (in reverse linkage disequilibrium with UCSNP-19) has previously been associated with decreasing insulin sensitivity (Elbein *et al.*, 2002). This indicates that there

may be a difference in the effects of this alteration in the two different populations. Further functional testing will have to be undertaken in order to elucidate the mechanism of this difference.

Table 5.22: Meta-analysis of genotypes at the UCSNP-56 locus for the black Southern African, Cuban and German cohorts

Model	Genotype Comparison	Cohort size (cases/controls)	OR (95% CI)		p-value
			Fixed effects	*Random effects*	
Recessive	1,1 vs 1,2 and 2,2	788/570	0.78 (0.58-1.05)	0.74 (0.27-1.98)	NS
Dominant	1,1 and 1,2 vs 2,2	788/570	0.75 (0.59-0.95)	0.81 (0.46-1.44)	NS
Additive	1,1 vs 1,2	480/386	0.86 (0.63-1.18)	0.78 (0.33-1.83)	NS
	1,2 vs 2,2	649/442	0.80 (0.62-1.04)	0.84 (0.55-1.29)	NS
	1,1 vs 2,2	447/312	0.62 (0.43-0.89)	0.66 (0.21-2.06)	NS

UCSNP-56: 1 = A allele; 2 = G allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; vs = versus; the shaded block indicates the evidence for association of the 1,1 genotype versus the 2,2 genotype towards a protective effect under the fixed effects model.

Meta-analysis of this locus within the various cohorts described resulted in a single association detected under fixed effects as presented in the shaded block in Table 5.22. The 1,1 haplotype is protective compared to the 2,2 haplotype. This must be driven by the strong association detected in the Southern African cohort as discussed in Section 5.4.4.3. Thus this significant association is hidden by the absence of this effect in the Cuban population as determined in Section 5.4.4.6. Conclusions regarding the utilisation of meta-analysis in the investigation of T2D susceptibility are presented in Section 6.4.

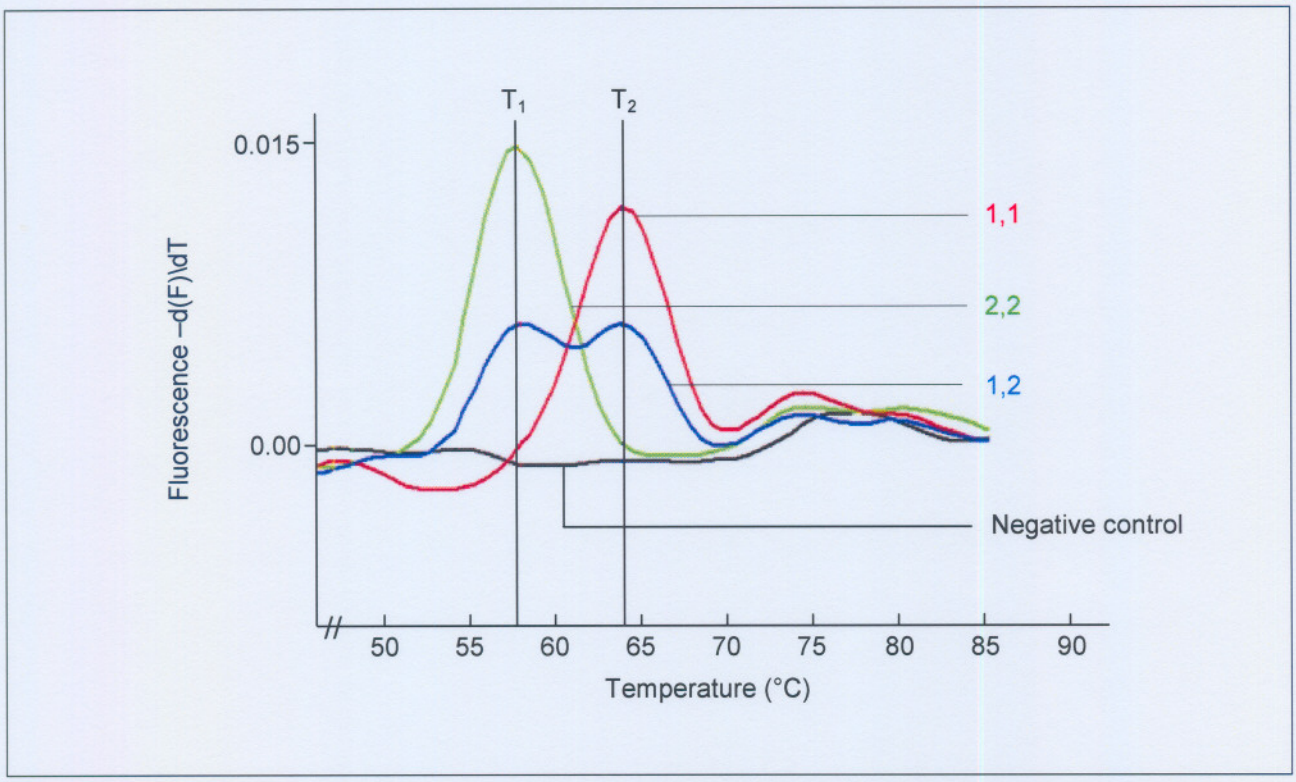
5.4.5 UCSNP-63 within the calpain 10 gene

Thermal cycling was achieved according to the protocol in Section 4.4. Amplification was optimised at 40 cycles of the thermal cycling parameters with a T_a of 64°C. Determination of the status of an individual at UCSNP-63 was achieved via manual inspection of the negative differential graph for fluorescence intensity determined from the melting curve as depicted in Figure 5.4.

The sensor probe of this alteration recognises the C allele and therefore the 1,1 genotype was assigned to the peak at the higher melting temperature (T_2 = 64.06°C, standard deviation = 1.43°C) as exemplified by the red curve in Figure 5.4, the 1,2 genotype to the curve with peaks at both melting temperatures, similar to the blue curve (heterozygote T_1 = 58.00°C, standard deviation = 1.56°C; heterozygote T_2 = 64.55°C, standard deviation = 1.30°C) and the 2,2 genotype to the peak at the lower temperature as is the

case for the green curve (T_1 = 57.70°C, standard deviation = 1.43°C). The 1,2 genotype was also only assigned if the two peaks of the curve generated were ca. half that of the respective curves generated for the 1,1 and 2,2 curves as discussed in Section 5.4.2. The peaks were calibrated according to the results detected for the two homozygote and the heterozygote controls.

Figure 5.4: Diagrammatic representation of the differential graph of probe fluorescence versus temperature for UCSNP-63 within the calpain 10 gene



UCSNP-63: 1 = C allele; 2 = T allele; 1,1 = homozygote for the 1 allele indicated by the red line; 1,2 = heterozygote indicated by the blue line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by black line wherein DNA was replaced by deionised water; T_1 = temperature at which the 2,2 homozygote peaks (T_1 = 57.70°C, standard deviation = 1.43°C); T_2 = temperature at which the 1,1 homozygote peaks (T_2 = 64.06°C, standard deviation = 1.43°C).

5.4.5.1 UCSNP-63 within the black Southern African diabetic cohort

The affected cohort under investigation was in HW equilibrium (χ^2 = 1.39) for UCSNP-63 as delineated in Table 5.23. The modifying effects described in Section 5.3.2 are therefore negligible in this cohort at this locus. The fact that the variant allele is present at such a high frequency may be indicative of association, however, comparison of the two cohorts is required before this can be hypothesised.

Table 5.23: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African patient cohort for UCSNP-63

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	52	103	70	225
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.23	0.46	0.31	1.00
Expected numbers (E)	47.61	111.78	65.61	225.00
$\chi^2 = (O-E)^2/E$	0.40	0.69	0.29	1.38

UCSNP-63: 1 = C allele; 2 = T allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.5.2 UCSNP-63 within the black Southern African control cohort

UCSNP-63 was analysed in the black South African control cohort and the locus was determined to be in HW equilibrium ($\chi^2 = 1.48$) as presented in Table 5.24. The various frequency modifying factors described in Section 5.3.2 do not have an effect at this locus. Any differences in genotype frequencies between the control and patient cohorts will indicate association between the specific genotype and T2D.

Table 5.24: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African control cohort for UCSNP-63

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	58	100	60	218
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.27	0.46	0.28	1.00
Expected numbers (E)	53.50	108.99	55.50	218.00
$\chi^2 = (O-E)^2/E$	0.38	0.74	0.36	1.48

UCSNP-63: 1 = C allele; 2 = T allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.5.3 Comparison of UCSNP-63 between both black Southern African cohorts

Although the variant allele at this locus is at a very high frequency, upon comparison of the Southern African control and diabetic cohorts, there was no significant difference in genotype distributions as indicated in Table 5.25. As discussed in Section 5.4.5.7 this allele is at a much lower frequency in various non-African populations (Horikawa *et al.*, 2000; Evans *et al.*, 2001; Cassell *et al.*, 2002). Thus there are definitely differences between populations at the level of genotype distribution.

Table 5.25: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-63 locus between the black Southern African patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.23 (n=52)	0.46 (n=103)	0.31 (n=70)	NS
Control	0.27 (n=58)	0.46 (n=100)	0.28 (n=60)	
OR (95%CI)	0.82 (0.53-1.27)	0.99 (0.68-1.44)	1.18 (0.78-1.79)	

UCSNP-63: 1 = C allele; 2 = T allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; n = sample size.

5.4.5.4 UCSNP-63 within the Cuban diabetic cohort

It was detected that genotypes at the UCSNP-63 locus were not in HW equilibrium (χ^2 = 11.34) within the Cuban patient cohort investigated as presented in Table 5.26. This is surprising due to the fact that the cohort is in HW equilibrium at the various other loci within the CAPN10 gene. There must therefore be a locus-specific effect affecting HW equilibrium. Generally an increase in homozygosity at the expense of heterozygosity indicates that inbreeding has taken place. However, this is unlikely as the other three SNPs are in fact in HW equilibrium. The second explanation for lower heterozygosity than expected is that the population sampled is erroneously assumed to be a single population when in fact it consists of several subpopulations. Each of these populations could well be in HW equilibrium, but due to the pooling effect of the sampling strategy, it ultimately results in lower levels of heterozygosity. However as this effect acts upon the entire locus, it would mean that the various other SNPs detected would also have a similar genotype profile. This, however, is not the case as discussed above.

Table 5.26: Chi-square test of goodness-of-fit to the HW proportions of the Cuban patient cohort for UCSNP-63

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	210	81	22	313
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.67	0.26	0.07	1.00
Expected numbers (E)	200.48	100.04	12.48	313.00
$\chi^2 = (O-E)^2/E$	0.45	3.62	7.26	11.34

UCSNP-63: 1 = C allele; 2 = T allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

In order to explain the lack of HW equilibrium, it must be taken into consideration that although this is a random sample within the selection conditions, the utilisation of these criteria causes the population investigated to be non-random in terms of the greater population. Thus the absence of HW equilibrium at a locus due to increased homozygosity

is the expected result if a hypothesis of linkage disequilibrium between the locus and the disease phenotype is assumed. Thus it can be theorised that the absence of HW equilibrium is evidence that UCSNP-63 is a more important risk factor in the Cuban population than the other three SNPs described. However, future studies will have to be undertaken in order to elucidate if this is in fact the case.

5.4.5.5 UCSNP-63 within the Cuban control cohort

The UCSNP-63 locus was not in HW equilibrium within the patient group investigated. However, this locus is in HW equilibrium ($\chi^2 = 0.01$) within the control cohort as presented in Table 5.27. This lends greater credence to the hypothesis that the lack of HW equilibrium within the patient cohort is due to natural selection. However caution must be exercised upon comparing the two cohorts at this locus, as it is no longer possible to rule out the effects of the various factors that affect HW equilibrium. It will thus not be possible to significantly associate this alteration with disease risk. It will, however, still be feasible to determine if there is a significant difference in genotype distributions between the two cohorts.

Table 5.27: Chi-square test of goodness-of-fit to the HW proportions of the Cuban control cohort for UCSNP-63

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	177	61	5	243
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.73	0.25	0.02	1.00
Expected numbers (E)	177.19	60.63	5.19	243.00
$\chi^2 = (O-E)^2/E$	0.00	0.00	0.01	0.01

UCSNP-63: 1 = C allele; 2 = T allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.5.6 Comparison of UCSNP-63 between both Cuban cohorts

The absence of HW equilibrium within the patient cohort at UCSNP-63 however complicates the comparison of the two cohorts for this locus. It is still possible to compare this cohort to the control cohort, which is in HW equilibrium, but any variation could be due to the numerous effects, which have been discussed in Section 5.3.2. The genotype frequencies at this locus were in fact significantly different upon comparison to the control cohort (p value < 0.0001) as presented in Table 5.28, specifically the variant allele in its homozygous state. This was determined upon calculation of an OR value of 3.59 (95% CI

1.34-9.64) presented in the shaded block in Table 5.28. This is however a rather wide CI and a larger cohort should thus be investigated in the future. It is not possible to definitively associate this allele with T2D as the significant difference determined may be due to a factor affecting HW equilibrium status. However if the previous arguments are taken into consideration it is possible that this link may be due to this locus undergoing natural selection. Further functional studies will have to be undertaken to determine the role that the variant allele at UCSNP-63 has in T2D.

Table 5.28: Chi-square analysis for the comparison of the calpain 10 genotype distribution at the UCSNP-63 locus between the Cuban patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.67 (n=210)	0.26 (n=81)	0.07 (n=22)	p<0.0001
Control	0.73 (n=177)	0.25 (n=61)	0.02 (n=5)	
OR (95% CI)	0.76 (0.52-1.09)	1.04 (0.70-1.53)	3.59 (1.34-9.64)	

UCSNP-63: 1 = C allele; 2 = T allele; OR = odds ratio; 95% CI = 95% confidence interval; n = sample size; the shaded block indicates the evidence for association of the 2,2 genotype with increased risk.

5.4.5.7 Meta-analysis of UCSNP-63

Upon analysis of the various loci within the Cuban cohorts investigated, the most important result detected was the absence of HW equilibrium at UCSNP-63. The fact that this alteration was in HW equilibrium within the control population further strengthens the evidence that this SNP may be involved in disease pathogenesis. Upon further statistical analyses it became apparent that this alteration was important in the T2D phenotype as it presented with an OR of 3.59 (95% CI 1.34-9.64) as depicted in Table 5.28. Care must, however, be taken with any conclusions drawn from this result due to the broad range of the 95% CI and secondly, due to the fact that the diabetic cohort is not in HW equilibrium. Any of the factors discussed in Section 5.3.2 could therefore explain the variation detected in genotype distributions.

UCSNP-63 was not associated with T2D risk within the African cohort however the variant allele was present in the South African cohorts at a much higher frequency (0.51) than that detected in the various non-African populations (circa 0.05) previously investigated (Horikawa *et al.*, 2000; Evans *et al.*, 2001; Cassell *et al.*, 2002). The variant allele at UCSNP-63 was always observed on a haplotype with the wild type allele at UCSNP-43 in Mexican American and European populations (Horikawa *et al.*, 2000). This may explain the apparent lack of the variant allele at UCSNP-43 in the Southern African cohort. As the

variant allele at UCSNP-63 is so prevalent in the two cohorts investigated (frequency = 0.54 in patients and 0.51 in controls) it may be that this association prevents the accumulation of the 2 allele at UCSNP-43. This assumption is however based on the current data in non-African populations (Horikawa *et al.*, 2000). This relationship should be investigated within the African population in order to assess the possible epistatic effect between these two alleles.

From this investigation it is possible to distinguish the Cuban, black South African and non-African populations from each other at the genotype level. The protective effect of UCSNP-56 within the black South African cohort indicates that the CAPN10 locus has varied effects in different populations. This varied pathogenesis is possibly due to the interaction of alleles at this locus with various other risk loci within the genome. This hypothesis is discussed in Section 6.1.

Table 5.29: Meta-analysis of genotypes at the UCSNP-63 locus for the black Southern African, Cuban and German cohorts

Model	Genotype Comparison	Cohort size (cases/controls)	OR (95% CI)		p-value
			Fixed effects	*Random effects*	
Recessive	1,1 vs 1,2 and 2,2	770/549	0.78 (0.60-1.01)	0.78 (0.60-1.01)	NS
Dominant	1,1 and 1,2 vs 2,2	770/549	0.72 (0.50-1.05)	0.60 (0.26-1.42)	NS
Additive	1,1 vs 1,2	676/483	0.85 (0.65-1.13)	0.85 (0.65-1.13)	NS
	1,2 vs 2,2	312/236	0.76 (0.51-1.14)	0.66 (0.28-1.55)	NS
	1,1 vs 2,2	552/379	0.63 (0.41-0.99)	0.56 (0.25-1.26)	NS

UCSNP-63: 1 = C allele; 2 = T allele; OR = odds ratio; 95% CI = 95% confidence interval; vs = versus; NS = non significant.

Although at the cohort level significant differences were detected in genotype distribution this does not translate to any significant OR values upon meta-analytical analysis as depicted in Table 5.29. No significant association could be differentiated at this level although the Cuban diabetic cohort was not in HW equilibrium. The fact that both the random effects and fixed effects models have allowed for the calculation of similar results leads to the hypothesis that the populations do not have differentiable levels of statistical heterogeneity i.e. the intra cohort variance is much greater than the inter cohort variance.

5.4.6 Haplotype frequencies at the calpain 10 locus

Previous investigations have elucidated the importance of specific haplotype and haplotype combinations within the CAPN10 gene in the genetic susceptibility towards T2D (Horikawa *et al.*, 2000; Malecki *et al.*, 2002; Cassell *et al.*, 2002). It was determined that

the 112/121 haplotype and the homozygous 121 haplotype both represent "at risk" haplotypes. The presence of these specific haplotypes was determined within the black South African and Cuban cohorts under investigation.

CAPN10 SNP haplotype assignment is often achieved utilising the excitation-maximisation (EM) algorithm (Long *et al.*, 1995). The haplotypes of the non-ambiguous genotype combinations were initially assigned and dependent on the frequency of these haplotypes it was possible to structure the relative haplotype distribution of the ambiguous genotype combinations (Fullerton *et al.*, 2002). However this process was made considerably simpler as certain haplotypes were absent or at such a low frequency as to be ignored within the group of individuals investigated. If a haplotype is rare within a cohort, such as the 222 haplotype in the Southern African cohort, in Section 5.4.6.1, it is unlikely that this phase is present for an ambiguous genotype combination. For example an individual heterozygous at all three loci investigated can produce the following haplotype combinations: 111/222; 112/221, 122/211 or 121/212. However the 222 haplotype is at a frequency < 0.05 therefore the first combination is unlikely. In the second example the 221 haplotype is present at < 0.05 therefore this combination is also unlikely. Within the last combination the 212 haplotype is not present at a frequency > 0.05. Finally, both the 122 and 211 haplotypes are present at a frequency of greater than 0.1 thus this haplotype combination is most likely. This use of the algorithm therefore produces a relative distribution of possible haplotypes within the cohort.

As stated in Section 5.4.4 it was decided that the UCSNP-19 allele would be replaced by the UCSNP-56 allele for haplotype determination. The concomitant alteration in haplotype naming has also been discussed and was utilised in the haplotype and haplotype combination comparison of the Southern African and Cuban cohorts.

5.4.6.1 Calpain 10 haplotype analysis within the black Southern African cohorts

Haplotype analysis as presented in Table 5.30 resulted in the calculation of a significant association between the 122 haplotype and disease risk (OR 1.59, 95% CI 1.22-2.08) as indicated in Row B of Table 5.30. Upon determination of haplotypes in the Mexican American, German and Finnish populations by Horikawa *et al.* (2000) it was found that only four of the eight possible haplotypes were detectable in the cohorts investigated in these studies. These haplotypes are indicated in bold within Table 5.30. The Southern African cohort presented with a similar trend in that these four haplotypes were all present

at a frequency greater than 0.05, whereas the 221, 222 and 212 haplotypes were not. The only difference however was the significant enrichment of the 112 haplotype in the control cohort. In turn, it was determined that a protective factor was in association with this haplotype (OR 0.14, 95% CI (0.06-0.31)) as depicted in Row D of Table 5.30.

Table 5.30: Chi-square analysis for comparison of the calpain 10 haplotype distribution between the black Southern African patient and control cohorts

	Haplotypes (43/56/63)	Patients Frequency (n)	Controls Frequency (n)	Chi square value (χ^2)	OR 95% CI	Haplotypes (43/19/63)
A	**121**	0.22 (98)	0.20 (86)	1.05	1.13 (0.82-1.57)	**111**
B	**122**	0.51 (228)	0.40 (172)	14.84	1.59 (1.22-2.08)	**112**
C	**111**	0.12 (55)	0.16 (70)	3.99	0.73 (0.49-1.07)	**121**
D	112	0.02 (7)	0.10 (44)	32.31	0.14 (0.06-0.31)	122
E	221	0.01 (3)	0.01 (4)	0.30	0.72 (0.16-3.27)	211
F	**211**	0.11 (48)	0.13 (56)	1.59	0.81 (0.53-1.22)	**221**
G	222	0.01 (5)	0.00 (0)	ND	ND	212
H	212	0.00 (0)	0.00 (0)	ND	ND	222
I	*Total*	*n = 444*	*n = 432*	*54.08*	---	---

$\chi^2 = 16.81$ for $\alpha = 0.01$ with df = 6; OR = odds ratio; CI = confidence interval; ND = non differentiable; constitution of haplotype indicated by blue text (xxx) = UCSNP-43, UCSNP-56, UCSNP-63; constitution of haplotypes indicated by green text (yyy) = UCSNP-43, UCSNP-19, UCSNP-63; values indicated by red text (zzz) indicate significant results discussed within the text; bold haplotypes indicate the only haplotypes present in the Mexican American, German and Finnish cohorts investigated by Horikawa *et al.* (2000); UCSNP-43: 1 = G allele; 2 = A allele; UCSNP-19: 1 = two repeat allele; 2 = three repeat allele; UCSNP-56: 1 = A allele; 2 = G allele; UCSNP-63: 1 = C allele; 2 = T allele; n = individuals for whom haplotypes, reflecting full genotypes at all the SNP loci, were generated.

Specific haplotypes at the CAPN10 locus have been significantly associated with disease risk in other populations (Horikawa *et al.*, 2000). The 211 haplotype for example, has yet to be detected in the Western Africans (Fullerton *et al.*, 2002). In contrast to this, the control population investigated in this report harbours this haplotype at a frequency of 0.13 while it is present in the affected population at a frequency of 0.11.

5.4.6.2 Calpain 10 haplotype combination analysis within the black Southern African cohorts

Analysis of the haplotype combinations as presented in Table 5.31 allowed for the determination that the non-African "at risk" 122/111 haplotype combination as indicated in bold in Row K was present at frequencies of 0.14 in the diabetic group and 0.12 in the control group. No association between the T2D phenotype and this haplotype combination could be detected in the cohorts investigated. The homozygous 111 haplotype combination which is presented in bold in Row L of Table 5.31 was present in less than 0.03 of the control cohort investigated and was deemed not to be a major risk factor for T2D in this population.

Table 5.31: Chi-square analysis for comparison of the calpain 10 haplotype combination distribution between the black Southern African patient and control cohorts

	Haplotype Combination (43/56/63)	Patients Frequency (n)	Controls Frequency (n)	Chi square value (χ^2)	OR 95% CI	Haplotype Combination (43/19/63)
A	121/121	0.06 (14)	0.05 (10)	1.35	1.38 (0.60-3.19)	111/111
B	121/122	0.19 (43)	0.14 (31)	3.89	1.43 (0.86-2.37)	111/112
C	121/111	0.08 (18)	0.08 (18)	0.01	0.97 (0.49-1.91)	111/121
D	121/221	0.00 (0)	0.01 (3)	3.08	ND	111/211
E	121/211	0.04 (9)	0.06 (14)	2.02	0.60 (0.25-1.43)	111/221
F	122/122	0.26 (58)	0.19 (40)	6.94	1.55 (0.98-2.45)	112/112
G	122/112	0.02 (5)	0.01 (1)	15.35	4.95 (0.57-42.7)	112/122
H	122/221	0.00 (0)	0.00 (0)	ND	ND	112/211
I	122/222	0.02 (5)	0.00 (0)	ND	ND	112/212
J	122/211	0.11 (24)	0.15 (33)	2.90	0.67 (0.38-1.18)	112/221
K	**122/111**	0.15 (32)	0.12 (26)	1.04	1.23 (0.70-2.14)	**112/121**
L	**111/111**	0.00 (0)	0.03 (6)	6.17	ND	**121/121**
M	111/112	0.00 (0)	0.04 (9)	9.25	ND	121/122
N	111/211	0.02 (5)	0.02 (5)	0.00	0.97 (0.27-3.40)	121/221
O	112/112	0.01 (1)	0.08 (17)	15.53	0.05 (0.00-0.40)	122/122
P	221/122	0.02 (3)	0.01 (1)	3.78	2.94 (0.30-28.5)	211/112
Q	221/211	0.00 (0)	0.00 (0)	ND	ND	211/221
R	221/221	0.00 (0)	0.00 (0)	ND	ND	211/211
S	211/211	0.02 (5)	0.01 (2)	4.22	2.46 (0.47-12.8)	221/221
T	*Total*	*n = 222*	*n = 216*	75.53	---	---

χ^2 = 27.69 for α = 0.01 with df = 13; OR = odds ratio; CI = confidence interval; ND = non differentiable; constitution of haplotype combination indicated by blue text (xxx) = UCSNP-43, UCSNP-56, UCSNP-63/ UCSNP-43, UCSNP-56, UCSNP-63; constitution of haplotypes indicated by green text (yyy) = UCSNP-43, UCSNP-19, UCSNP-63/ UCSNP-43, UCSNP-19, UCSNP-63; values indicated by red text (zzz) indicate significant results discussed within the text; bold haplotype combinations indicate the "at risk" haplotype combinations detected in the Mexican American, German and Finnish populations (Horikawa *et al.*, 2000) and the Polish population (Malecki *et al.*, 2002) respectively; UCSNP-43: 1 = G allele; 2 = A allele; UCSNP-19: 1 = two repeat allele; 2 = three repeat allele; UCSNP-56: 1 = A allele; 2 = G allele; UCSNP-63: 1 = C allele; 2 = T allele; n = individuals for whom haplotype combinations, reflecting full genotypes at all the SNP loci, were generated.

The 112/112 haplotype combination was significantly associated (p-value < 0.0001) with a protective effect (OR 0.05, 95% CI (0.00-0.40)) as presented in Row O of Table 5.31. The 122/122 haplotype alternatively, presented with marginal significance (p < 0.01) to increased risk as indicated by an OR of 1.55 as depicted in Row F of Table 5.31. The 95% CI does however span values both below and above 1.0 which, as discussed in Section 5.3, is indicative of no increased or decreased risk. Sampling a larger cohort may narrow the interval to exclude values below one but care should be taken in conclusions determined from such a value. If this association is indeed true it increases the evidence

for the association found at the haplotype level discussed in Section 5.4.4.3. As the only differentiable aspect of these two haplotype combinations is the allele present at UCSNP-56 it may be that in the absence of this allele the 122 haplotype increases disease risk, however the presence of the wild type allele protects against this effect thus indicating that UCSNP-56 is a significant determinant of disease risk.

There is a large chi-square value determined for the comparison of the various haplotype combinations within the cohorts investigated. This is due to the individuals harbouring the 122/112 haplotype combination (OR 4.95, 95% CI (0.57-42.7)) highlighted in red in Row G of Table 5.31. Although there is a significant difference determined via χ^2 analysis, upon determination of the OR, the 95% CI was too wide due to the relatively low number of individuals harbouring this specific combination. Larger cohorts will have to be investigated in order to definitively determine the role this haplotype has on T2D pathogenesis.

5.4.6.3 Calpain 10 haplotype analysis within the Cuban cohorts

The utilisation of χ^2 analysis as depicted in Table 5.32 allowed for the determination that the 122 haplotype was significantly associated with increased disease risk (χ^2 = 9.70; OR 1.40, 95% CI 1.01-1.94) as highlighted by the red text in Row B of Table 5.32. This association is however marginal and further analysis is required to determine the true significance of this haplotype.

Table 5.32: Chi-square analysis for comparison of the calpain 10 haplotype distribution between the Cuban patient and control cohorts

	Haplotypes (43/56/63)	Patients Frequency (n)	Controls Frequency (n)	Chi square value (χ^2)	OR 95% CI	Haplotypes (43/19/63)
A	**121**	0.27 (167)	0.27 (132)	0.04	0.97 (0.74-1.27)	**111**
B	**122**	0.19 (118)	0.14 (69)	9.70	1.40 (1.01-1.94)	**112**
C	**111**	0.33 (203)	0.35 (169)	0.93	0.90 (0.70-1.16)	**121**
D	112	0.01 (4)	0.00 (1)	5.74	3.12 (0.34-28.0)	122
E	221	0.01 (7)	0.01 (6)	0.07	0.90 (0.30-2.71)	211
F	**211**	0.20 (123)	0.22 (106)	1.28	0.87 (0.65-1.17)	**221**
G	222	0.00 (0)	0.00 (1)	1.29	ND	212
H	212	0.00 (0)	0.00 (0)	ND	ND	222
I	*Total*	*n = 622*	*n = 484*	*17.75*	---	---

χ^2 = 16.81 for α = 0.01 with df = 6; OR = odds ratio; CI = confidence interval; ND = non differentiable; constitution of haplotype indicated by blue text (xxx) = UCSNP-43, UCSNP-56, UCSNP-63; constitution of haplotypes indicated by green text (yyy) = UCSNP-43, UCSNP-19, UCSNP-63; values indicated by red text (zzz) indicate significant results discussed within the text; bold haplotypes indicate the only haplotypes present in the Mexican American, German and Finnish cohorts investigated by Horikawa *et al.* (2000); UCSNP-43: 1 = G allele; 2 = A allele; UCSNP-19: 1 = two repeat allele; 2 = three repeat allele; UCSNP-56: 1 = A allele; 2 = G allele; UCSNP-63: 1 = C allele; 2 = T allele; n = individuals for whom haplotypes, reflecting full genotypes at all the SNP loci, were generated.

As in the discussion in Section 5.4.6.1, the Cuban cohort also only harbours four of the possible eight haplotypes at an appreciable level as indicated by the bold text in Table 5.32. Similar to the various non-African populations (Fullerton et al., 2002) the 111 haplotype is the most prevalent in the Cuban cohorts investigated. It has a frequency of 0.33 in the patient group and 0.35 in the control group. It does not however present with any association to T2D.

No other haplotypes presented with any significant association within the Cuban cohorts investigated. Although the 112 haplotype presented in Row D of Table 5.32, has a high chi-square value, this haplotype was only present in four patients and one control individual. The confidence interval of the odds ratio calculation is therefore extremely wide thus indicating that there is a high probability of a type I error. Thus it will be necessary to investigate a larger cohort to determine if this association is maintained.

5.4.6.4 Calpain 10 haplotype combination analysis within the Cuban cohorts

As no further associations could be determined with the haplotypes, the various haplotype combinations were investigated utilising a similar strategy as presented in Table 5.31. The results of these analyses are presented in Table 5.33. The most significant association was determined for the homozygous 122 haplotype combination. As this haplotype presented with association as discussed in Section 5.5.1.2 it is not surprising that this association is maintained at the level of haplotype combination.

As discussed in Section 3.3.1.5 the 122/111 and 111/111 haplotype combinations are associated with disease risk in various non-African populations (Horikawa *et al.*, 2000; Malecki *et al.*, 2002). These combinations do not present with any significant association in the Cuban cohorts analysed as depicted in Rows K and L in Table 5.33 respectively. The 122/122 haplotype combination on the other hand presents with an OR of 4.30 (95%CI 1.45-12.7) as presented in Row F of Table 5.33. This is similar to the South African cohort investigated however the 95% CI is extremely wide thus increasing the risk of a type 1 error and hence further analyses are required before this risk allele is definitive.

Table 5.33: Chi-square analysis for comparison of the calpain 10 haplotype combination distribution between the Cuban patient and control cohorts

	Haplotype Combination (43/56/63)	Patients Frequency (n)	Controls Frequency (n)	Chi square value (χ^2)	OR 95% CI	Haplotype Combination (43/19/63)
A	121/121	0.08 (26)	0.08 (19)	0.10	1.07 (0.57-1.98)	111/111
B	121/122	0.08 (25)	0.07 (16)	0.96	1.23 (0.64-2.36)	111/112
C	121/111	0.17 (53)	0.21 (51)	2.40	0.76 (0.50-1.17)	111/121
D	121/221	0.01 (2)	0.01 (2)	0.13	0.77 (0.10-5.55)	111/211
E	121/211	0.11 (35)	0.10 (25)	0.26	1.10 (0.63-1.89)	111/221
F	122/122	0.07 (21)	0.02 (4)	48.93	4.30 (1.45-12.7)	112/112
G	122/112	0.00 (0)	0.00 (0)	ND	ND	112/122
H	122/221	0.00 (0)	0.01 (3)	3.86	ND	112/122
I	122/222	0.00 (0)	0.00 (1)	1.29	ND	112/212
J	122/211	0.06 (18)	0.10 (23)	4.52	0.58 (0.30-1.11)	112/221
K	**122/111**	0.11 (33)	0.07 (18)	4.21	1.47 (0.81-2.69)	**112/121**
L	**111/111**	0.11 (35)	0.12 (30)	0.33	0.89 (0.53-1.50)	**121/121**
M	111/112	0.01 (4)	0.00 (1)	5.74	3.14 (0.34-28.2)	121/122
N	111/211	0.14 (43)	0.16 (39)	1.01	0.83 (0.52-1.33)	121/221
O	112/112	0.00 (0)	0.00 (0)	ND	ND	122/122
P	221/122	0.00 (0)	0.00 (0)	ND	ND	211/112
Q	221/211	0.00 (1)	0.00 (1)	0.06	0.77 (0.04-12.4)	211/221
R	221/221	0.01 (2)	0.00 (0)	ND	ND	211/211
S	211/211	0.04 (13)	0.04 (9)	0.18	1.12 (0.47-2.68)	221/221
T	*Total*	*n = 311*	*n = 242*	73.96	---	

χ^2 = 27.69 for α = 0.01 with df = 13; OR = odds ratio; CI = confidence interval; ND = non differentiable; constitution of haplotype combination indicated by blue text (xxx) = UCSNP-43, UCSNP-56, UCSNP-63/ UCSNP-43, UCSNP-56, UCSNP-63; constitution of haplotypes indicated by green text (yyy) = UCSNP-43, UCSNP-19, UCSNP-63/ UCSNP-43, UCSNP-19, UCSNP-63; values indicated by red text (zzz) indicate significant results discussed within the text; bold haplotype combinations indicate the "at risk" haplotype combinations detected in the Mexican American, German and Finnish populations (Horikawa *et al.*, 2000) and the Polish population (Malecki *et al.* 2002) respectively; UCSNP-43: 1 = G allele; 2 = A allele; UCSNP-19: 1 = two repeat allele; 2 = three repeat allele; UCSNP-56: 1 = A allele; 2 = G allele; UCSNP-63: 1 = C allele; 2 = T allele; n = individuals for whom haplotype combinations, reflecting full genotypes at all the SNP loci, were generated.

5.4.6.5 Comparison of calpain 10 haplotypes and haplotype combinations to non-African populations

It was determined that the most prevalent calpain 10 haplotype structure within the various non-African populations investigated was that of the 111 haplotype, which was not present at a frequency greater than 0.05 in the African populations previously investigated (Fullerton *et al.*, 2002). In contrast, the frequency of this haplotype is 0.12 in the affected individuals and 0.16 in the control individuals investigated. The reason for this apparent incongruence may be the fact that the population groups previously investigated consisted of smaller cohorts for the Biaka (n = 138) and Mbuti (n = 70) pygmies (Fullerton *et al.*,

2002). This investigation consisted of n = 222 and n = 216 for the affected and unaffected cohorts respectively. The low frequencies of the previously discussed haplotypes may be attributable to selection bias due to the small sample size in the earlier study.

Although the sample size of the current investigation may have contributed to the under representation of the low frequency haplotypes, it does however garner greater evidence for the fact that the 122 haplotype is possibly the most prevalent in the African population. This haplotype was detected at a frequency of approximately 0.51 and 0.4 in the affected and control cohorts. If it is taken into consideration that this haplotype is significantly associated with disease risk its enrichment is rather surprising. As previously discussed the various non-African populations arose via migratory events from the African population (Cann *et al.*, 1987). It is speculated that this could explain the decrease in the levels of the 122 haplotype in the non-Africans when compared to the Africans as it may have been caused by an adaptive advantage in the African population which was lost upon migration out of this land mass.

At the level of haplotype combination, as presented in Table 5.31, the 122/111 non-African risk haplotype combination does not present with any significant association to T2D in the South African cohort. Furthermore the 111/111 haplotype combination, the other non-African risk factor, is absent from the patient population investigated thus indicating that it is unlikely to be a risk factor in this case. There is, however, evidence of a protective factor due to the increased level of the 111/111 haplotype combination in the control cohort. Unfortunately the odds ratio value was non-differentiable due to the haplotype combination being absent in the black South African patient cohort.

There are multiple reasons for the seeming absence of association with the various non-African "at risk" genetic factors. The most obvious may be that CAPN10 may not be a relevant susceptibility locus for T2D in the Southern African or Cuban populations. This is a rather simplistic view since there are alleles within this gene that are strongly associated with the disease risk i.e. the wild type allele of UCSNP-56.

A more likely explanation of the observed results is that different genetic factors within CAPN10 are responsible for disease risk in the black Southern African and Cuban populations. The complex nature of T2D suggests that there are loci which increase disease susceptibility and those which decrease it (Cox, 2001). Thus these pathways are more interactive than the loci involved in monogenic disease allowing for natural selection to have a more subtle effect. In addition the non-African populations have over time been

subjected to different selective and adaptive forces such as climate (Mishmar *et al.*, 2003), which may have allowed for the enrichment of distinctive genotypes within these groups.

It therefore becomes apparent that in developing strategies for the treatment of T2D it is necessary to take the genetic background, including ancestral lineages of an individual or population, into consideration. Due to the small effects that each of these genetic factors may have on disease susceptibility, it is not surprising that their effects are not detectable upon investigation of large multi-ethnic cohorts. The results presented thus support the elucidation of T2D susceptibility loci in a population dependent manner.

5.4.7 Detection of single nucleotide polymorphisms within the adiponectin gene

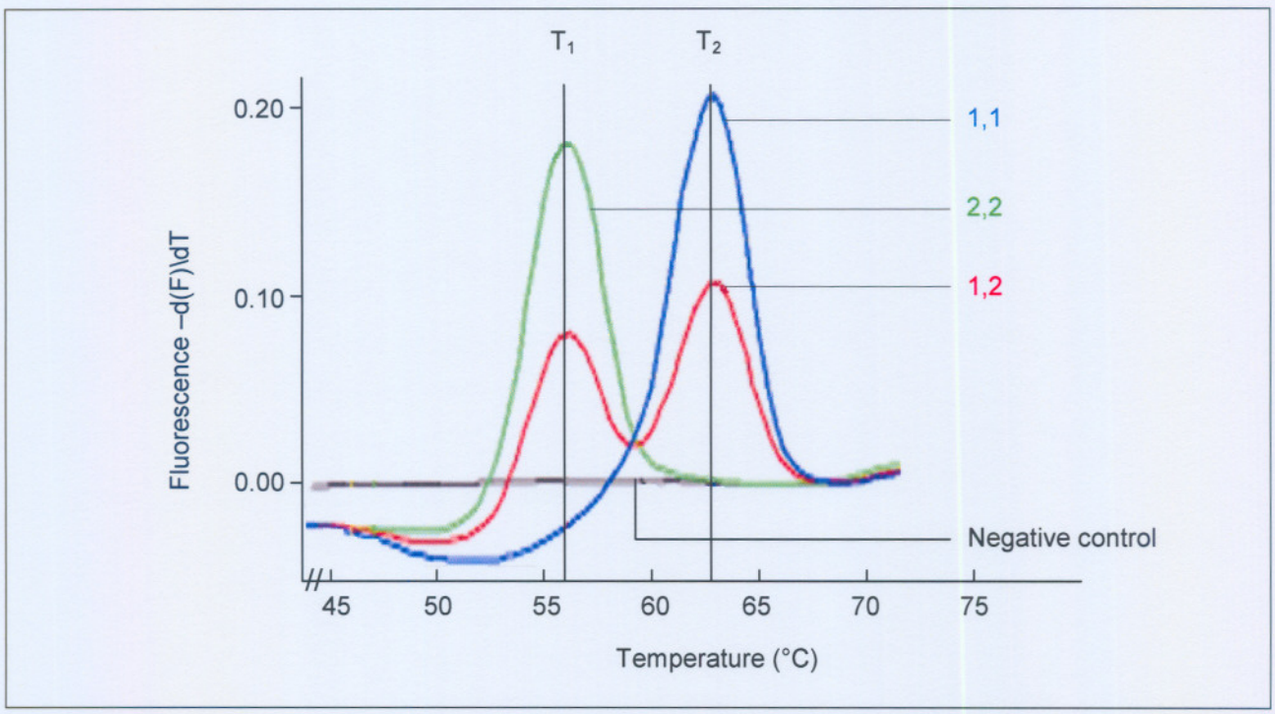
The APM1 gene has undergone numerous genetic investigations to determine association with T2D (Vasseur *et al.*, 2002; Hara *et al.*, 2002). Due to the associations described it was decided that the two alterations namely C-11377G and T45G within the APM1 gene would be analysed within the black South African and Cuban cohorts according to the methods discussed in Section 4.3 and 4.4. Although not an original aim, the G-11391A alteration was also screened in the black Southern African cohorts in order to investigate if this alteration is associated with disease risk. Furthermore it allowed for the generation of haplotypes which were comparable to the investigation by Schwarz *et al.* (2004). The results of these analyses are discussed in the following sections.

As stated earlier alterations in the adiponectin gene are associated with T2D risk and increased BMI. The importance of family history to T2D risk has been determined in the South African population however very little is known with regard to the genetic factors that are at play (Erasmus *et al.*, 2001). It was determined that the majority of female diabetics in the African population tend to present with high BMI values (Omar *et al.*, 1993), leading to the hypothesis that the APM1 gene may play a significant role. As mentioned previously in Section 5.1.1.1 the black African population belongs to macrohaplogroup L that harbours the most ancient human lineage as well as having the highest level of variation (Chen *et al.*, 1995). The high level of genetic variation may be synchronous to the clinical heterogeneity observed in the African diabetic population. This investigation was undertaken to determine if the genetic structures within the APM1 gene that are associated with disease risk in the Caucasian population, were similarly associated in the African and Cuban populations.

5.4.8 C-11377G within the adiponectin gene

This alteration was screened via mutation detection analysis of the promoter region of the APM1 gene. Amplification of this region of the gene was achieved via 35 cycles of the cycling parameters described in Table 4.6 with a T_a of 58°C. The peaks produced for this alteration upon investigation of the negative differential graph of the melting curve, were calibrated according to the peaks determined for the three positive controls as displayed in Figure 5.5.

Figure 5.5: Diagrammatic representation of the differential graph of probe fluorescence versus temperature for C-11377G within the adiponectin gene



Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; 1,1 = homozygote for the 1 allele indicated by the blue line; 1,2 = heterozygote indicated by the red line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by black line wherein DNA was replaced by deionised water; T_1 = temperature at which the 2,2 homozygote peaks (T_1 = 55.99°C, standard deviation = 1.17°C); T_2 = temperature at which the 1,1 homozygote peaks (T_2 = 62.71°C, standard deviation = 0.94°C).

The 1,1 genotype was only assigned to individuals with a peak at the higher melting temperature as is the case for the blue curve in Figure 5.5 (T_2 = 62.71°C, standard deviation = 0.94°C), as the sensor probe specifically recognises the C allele. Therefore the peak of a sample at the lower melting temperature (T_1 = 55.99°C, standard deviation = 1.17°C) was designated a 2,2 genotype as indicated by the green curve while the heterozygote was scored if the graph had two peaks present at both the melting temperatures similar to the red curve (T_1 = 56.04°C, standard deviation = 1.39°C; T_2 = 62.99°C, standard deviation = 0.97°C). In the case of two peaks being produced as indicated by the red curve, it was necessary to ensure that the amplitude of the two peaks

were approximately half that of the respective homozygote peaks. This is discussed in greater detail in Section 5.4.2.

5.4.8.1 Adiponectin SNP C-11377G within the black Southern African diabetic cohort

The C-11377G alteration was calculated to be in HW equilibrium in the patient cohort under investigation ($\chi^2 = 3.66$) as presented in Table 5.34. The patient cohort thus does not present with any alterations in genotype frequencies due to the factors described in Section 5.3.2. Any significant alteration in the genotype frequencies between this and the control cohort is indicative of association with disease risk.

Table 5.34: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African patient cohort for C-11377G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	171	63	1	235
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.73	0.27	0.00	1.00
Expected numbers (E)	174.49	56.01	4.49	235.00
$\chi^2 = (O-E)^2/E$	0.07	0.87	2.72	3.66

Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.8.2 Adiponectin SNP C-11377G within the black Southern African control cohort

Analysis of the C-11377G alteration within the control cohort has determined that it is experiencing HW equilibrium ($\chi^2 = 0.04$) as depicted in Table 5.35. This therefore allowed for the comparison of the control cohort to the patient cohort. Any significant difference in genotype frequencies between the two groups could thus not be attributed to the effects discussed in Section 5.3.2.

Table 5.35: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African control cohort for C-11377G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	166	61	5	232
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.72	0.26	0.02	1.00
Expected numbers (E)	166.43	60.14	5.43	232.00
$\chi^2 = (O-E)^2/E$	0.00	0.01	0.03	0.04

Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.8.3 Comparison of adiponectin SNP C-11377G between both black Southern African cohorts

No significant difference was present between the Southern African diabetic and control cohorts upon χ^2 analysis, as presented in Table 5.36. This allows for the hypothesis that C-11377G is not a significant risk factor within the Southern African population. The low OR ratio for the homozygote of the variant allele may be indicative of a possible association of this genotype and a protective factor against T2D susceptibility however as it is at a relatively low frequency within these cohorts (frequency = 0.00 and 0.02 in the diabetic and control cohorts respectively) it may ultimately be determined not to be a significant factor in mitigating disease risk.

Table 5.36: Chi-square analysis for the comparison of the adiponectin genotype distribution at the C-11377G locus between the black Southern African patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.73 (n=171)	0.27 (n=63)	0.00 (n=1)	NS
Control	0.72 (n=166)	0.26 (n=61)	0.02 (n=5)	
OR (95% CI)	1.06 (0.70-1.59)	1.02 (0.68-1.54)	0.19 (0.02-1.67)	

Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; χ^2 = 9.21 for α = 0.01 with df = 2; OR = odds ratio; NS = non significant; n = sample size.

5.4.8.4 Adiponectin SNP C-11377G within the Cuban diabetic cohort

The genotype distribution for C-11377G within the Cuban cohort is presented in Table 5.37. The cohort is in HW equilibrium at this locus (χ^2 = 0.79). It may therefore be hypothesised that the assumptions of HW equilibrium as discussed in Section 5.3.2 may be applied to this cohort.

Table 5.37: Chi-square test of goodness-of-fit to the HW proportions of the Cuban patient cohort for C-11377G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	198	107	19	324
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.61	0.33	0.06	1.00
Expected numbers (E)	195.22	112.55	16.22	324.00
$\chi^2 = (O-E)^2/E$	0.04	0.27	0.48	0.79

Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.8.5 Adiponectin SNP C-11377G within the Cuban control cohort

Investigation of the genotype distribution of SNP C-11377G within the Cuban control cohort allowed for the determination that it is currently in HW equilibrium ($\chi^2 = 7.27$) at this locus. The genotypic frequencies are presented in Table 5.38. This cohort is therefore not affected by the various factors discussed in Section 5.3.2.

Table 5.38: Chi-square test of goodness-of-fit to the HW proportions of the Cuban control cohort for C-11377G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	140	105	6	251
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.56	0.42	0.02	1.00
Expected numbers (E)	147.63	89.73	13.63	251.00
$\chi^2 = (O-E)^2/E$	0.39	2.60	4.27	7.27

Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.8.6 Comparison of adiponectin SNP C-11377G between both Cuban cohorts

A significant difference was determined upon comparison of the genotype distributions of the two cohorts at the C-11377G locus (p value < 0.0001) as presented in Table 5.39. The homozygote of the variant allele was in association with increased risk due to the high OR value calculated (OR 2.54, 95% CI 1.00-6.45) as highlighted in the darker shaded block within Table 5.39. The 95% CI spans the 1.0 value and it is therefore possible that this association is spurious. The relatively wide interval may be due to the fact that the homozygote for the variant allele is relatively rare in the cohorts investigated (frequency = 0.06 and 0.02 in the patient and control cohorts respectively). It will be useful in the future to investigate a larger cohort in order to narrow the interval.

Conversely the 1,2 genotype was associated with a protective factor due to the calculated OR (OR 0.68, 95% CI 0.48-0.96) as depicted in the lighter shaded block within Table 5.39. The higher levels of this genotype in the control cohort may be due to heterozygote advantage caused by its association with a protective factor. There are numerous examples of heterozygote advantage such as the protection that an individual heterozygous at the locus for sickle cell anaemia has against malaria (Flint *et al.*, 1993). Functional analysis is required in the future in order to determine the exact nature of this association.

Table 5.39: Chi-square analysis for the comparison of the adiponectin genotype distribution at the C-11377G locus between the Cuban patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.61 (n=198)	0.33 (n=107)	0.06 (n=19)	<0.0001
Control	0.56 (n=140)	0.42 (n=105)	0.02 (n=6)	
OR (95% CI)	1.24 (0.89-1.74)	0.68 (0.48-0.96)	2.54 (1.00-6.46)	

Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; χ^2 = 9.21 for α = 0.01 with df = 2; OR = odds ratio; n = sample size; the lighter shaded block indicates the evidence for association of the 1,2 genotype with a protective effect; the darker shaded block indicates the evidence for association of the 2,2 genotype to increased disease risk.

5.4.8.7 Meta-analysis of adiponectin SNP C-11377G

French Caucasian diabetic cohorts harbour the variant alleles at this locus at a frequency of 0.29 (Vasseur *et al.*, 2002). As depicted in Table 5.36 the black South African diabetic cohort harboured these alterations at frequencies of 0.15 respectively, thus indicating the comparative rarity of the variant allele. The Cuban cohorts alternatively harboured the variant allele at the C-11377G locus at a frequency of 0.23 in the diabetic cohort.

Table 5.40: Meta-analysis of genotypes at the C-11377G locus for the black Southern African, Cuban and German cohorts

Model	Genotype Comparison	Cohort size (cases/controls)	OR (95% CI)		p-value
			Fixed effects	*Random effects*	
Recessive	1,1 vs 1,2 and 2,2	924/806	1.00 (0.82-1.21)	1.01 (0.77-1.33)	NS
Dominant	1,1 and 1,2 vs 2,2	924/806	0.61 (0.38-0.98)	0.70 (0.28-1.72)	NS
Additive	1,1 vs 1,2	869/775	1.05 (0.86-1.29)	1.06 (0.79-1.42)	NS
	1,2 vs 2,2	368/317	0.62 (0.38-1.02)	0.71 (0.25-1.97)	NS
	1,1 vs 2,2	611/520	0.61 (0.37-0.99)	0.70 (0.30-1.65)	NS

Adiponectin SNP C-11377G: 1 = C allele; 2 = G allele; OR = odds ratio; 95% CI = 95% confidence interval; NS = non significant; vs = versus; the lighter shaded block indicates the evidence for association of the 1,1 and 1,2 genotypes versus the 2,2 genotype with a protective effect under the fixed effects model; the darker shaded blocks indicate the evidence for association of the 1,1 and 1,2 genotypes versus the 2,2 genotype with a protective effect under the additive model respectively.

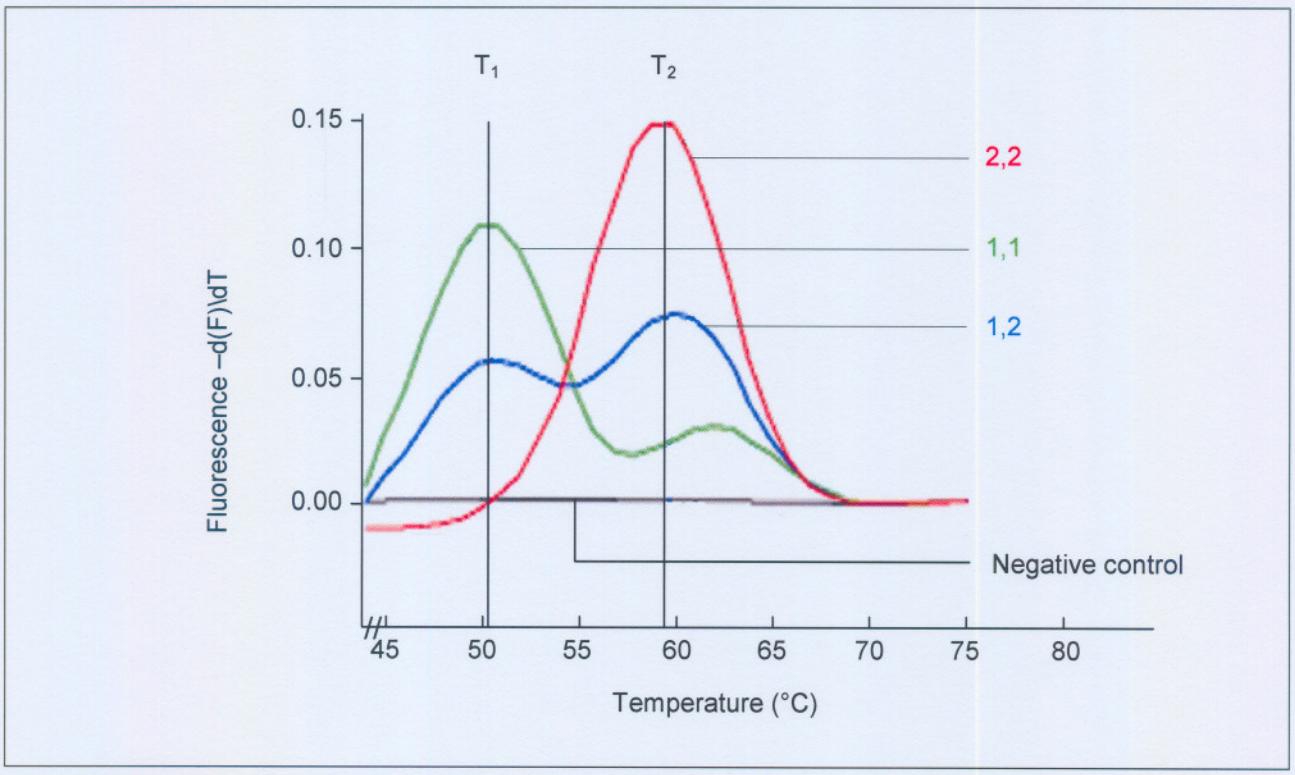
Meta-analytical investigation of the C-11377G locus within the black Southern African, Cuban and German populations resulted in the elucidation of a possible global association of the 1 allele with protection against disease, in the context of fixed effects as presented in the lighter shaded block in Table 5.40. This is due to the fact that under the dominant model the OR of 0.61 (95% CI 0.38-0.98) indicates that individuals harbouring the 1,1 and 1,2 genotypes are protected against disease risk as compared to homozygotes of the variant allele.

This is in agreement with the Cuban cohort analysis in Section 5.4.8.6 where it was determined that the heterozygote was associated with protection towards T2D susceptibility. This effect is maintained albeit marginally upon investigation of the additive models where the 2,2 homozygote is compared with the other two genotypes as indicated in the darker shaded block in Table 5.29. However this increase in risk is not significant upon calculation under random effects indicating that the inter population variance has an effect on the association of this possible risk factor. Ultimately it will be required in future studies to investigate the functional role of this polymorphism in these specific populations to understand its molecular pathogenesis in disease susceptibility.

5.4.9 T45G within the adiponectin gene

The region within the adiponectin gene harbouring this alteration was amplified via the thermal cycling template presented in Section 4.3. Amplification was optimised at 35 cycles with a T_a of 55°C. The melting curve was in turn utilised in the determination of the negative differential graph by which the various possible genotypes were assigned as indicated in Figure 5.6.

Figure 5.6: Diagrammatic representation of the differential graph of probe fluorescence versus temperature for T45G within the adiponectin gene



Adiponectin SNP T45G: 1 = T allele; 2 = G allele; 1,1 = homozygote for the 1 allele indicated by the red line; 1,2 = heterozygote indicated by the blue line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by black line wherein DNA was replaced by deionised water; T_1 = temperature at which the 1,1 homozygote peaks (T_1 = 50.14 °C, standard deviation = 0.69 °C); T_2 = temperature at which the 2,2 homozygote peaks (T_2 = 59.52 °C, standard deviation = 0.57 °C).

As the probe is specific for the variant allele, recognition of the 2 allele by the sensor probe resulted in the single peak at the higher temperature (T_2 = 59.52°C, standard deviation = 0.57°C) being indicative of a sample harbouring the 2,2 genotype as represented by the red curve in Figure 5.6. A single peak at the lower temperature (T_1 = 50.14°C, standard deviation = 0.69°C) was scored as a 1,1 homozygote (green curve) whereas the heterozygote was represented by two peaks at both temperatures similar to the blue curve (heterozygote T_1 = 49.88°C, standard deviation = 0.81°C; heterozygote T_2 = 59.79°C, standard deviation = 0.54°C), respectively. Similar to the discussion in Section 5.4.2 the amplitudes of the two peaks of the heterozygote curve had to be ca. half the amplitude of the relevant homozygote curves. Peaks were calibrated according to the graphs produced for the three positive controls.

5.4.9.1 Adiponectin SNP T45G within the black Southern African diabetic cohort

The heterozygotes and mutant homozygotes had to be grouped in order to investigate the HW equilibrium status of the cohort for T45G. It was determined that the group was in HW equilibrium at this locus (χ^2 = 0.00) as listed in Table 5.41 thus indicating that the assumptions discussed in Section 5.3.2 can be accepted with regard to this cohort.

Table 5.41: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African patient cohort for T45G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	228	4	0	232
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.98	0.02	0.00	1.00
Expected numbers (E)	228.02	3.97	Not calculated	232.00
$\chi^2 = (O-E)^2/E$	0.00	0.00	Not calculated	0.00

Adiponectin SNP T45G: 1 = T allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.9.2 Adiponectin SNP T45G within the black Southern African control cohort

As depicted in Table 5.42, HW equilibrium testing at the T45G locus resulted in the determination that this locus was in fact unaffected by such effects of population size and dynamics as inbreeding and migration (χ^2 = 0.04) as explained in Section 5.3.2. Effects at the level of the locus such as allele reversion, natural selection and random genetic drift

are also not responsible for the current genotype distribution observed. Therefore comparison of the patient and control cohorts for T45G is possible.

Table 5.42: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African control cohort for T45G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	220	6	0	226
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.97	0.03	0.00	1.00
Expected numbers (E)	220.04	5.92	0.04	226.00
$\chi^2 = (O-E)^2/E$	0.00	0.00	0.04	0.04

Adiponectin SNP T45G: 1 = T allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.9.3 Comparison of adiponectin SNP T45G between both black Southern African cohorts

The black Southern African cohort did not present with any significant association between any genotype at this locus and disease risk, as depicted in Table 5.43. This is however not surprising as the variant allele was present only at low levels (frequency of 0.01 and 0.02 in the diabetic and control cohorts respectively). Investigation of a larger cohort may narrow the 95% CI that was generated but ultimately this polymorphism may not be a risk allele in this population due to its rarity.

Table 5.43: Chi-square analysis for the comparison of the adiponectin genotype distribution at the T45G locus between the black Southern African patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.98 (n=228)	0.02 (n=4)	0.00	NS
Control	0.97 (n=220)	0.03 (n=6)	0.00	
OR (95% CI)	1.55 (0.43-5.58)	0.48 (0.11-1.94)	Not calculated	

Adiponectin SNP T45G: 1 = T allele; 2 = G allele; 2,2 homozygous carriers of the variant allele at T45G were not detected upon analyses of the various cohorts and therefore the sum of homozygotes for the variant allele and heterozygotes were utilised for statistical analyses to prevent non-differentiation due to division by zero; OR = odds ratio; NS = non significant; n = sample size.

5.4.9.4 Adiponectin SNP T45G within the Cuban diabetic cohort

As depicted in Table 5.44 the Cuban patient cohort is in HW equilibrium ($\chi^2 = 5.64$) at the T45G locus. Thus the effects described in Section 5.3.2 can be ignored. Furthermore the cohort may be utilised in the statistical determination of association with T2D susceptibility.

Table 5.44: Chi-square test of goodness-of-fit to the HW proportions of the Cuban patient cohort for T45G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	248	62	10	320
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.78	0.19	0.03	1.00
Expected numbers (E)	243.25	71.49	5.25	320.00
$\chi^2 = (O-E)^2/E$	0.09	1.26	4.29	5.64

Adiponectin SNP T45G: 1 = T allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.9.5 Adiponectin SNP T45G within the Cuban control cohort

The Cuban control cohort is not subjected to any of the factors discussed in Section 5.3.2. This is due to the fact that the T45G locus is in HW equilibrium ($\chi^2 = 3.27$) within this cohort. The genotypic frequencies as well as the calculation of HW status are indicated in Table 5.45.

Table 5.45: Chi-square test of goodness-of-fit to the HW proportions of the Cuban control cohort for T45G

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	188	56	9	253
Expected proportions	P^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.74	0.22	0.04	1.00
Expected numbers (E)	184.41	63.18	5.41	253.00
$\chi^2 = (O-E)^2/E$	0.07	0.82	2.38	3.27

Adiponectin SNP T45G: 1 = T allele; 2 = G allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.9.6 Comparison of adiponectin SNP T45G within both Cuban cohorts

None of the genotypes present at the T45G locus were determined to be significantly different between the Cuban cohorts investigated as presented in Table 5.46. It is therefore unlikely that this alteration is associated with disease risk within this population however future analysis may allow for the determination of a functional role for this variant.

Table 5.46: Chi-square analysis for the comparison of the adiponectin genotype distribution at the T45G locus between the Cuban patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.78 (n=248)	0.19 (n=62)	0.03 (n=10)	NS
Control	0.74 (n=188)	0.22 (n=56)	0.04 (n=9)	
OR (95% CI)	1.19 (0.81-1.75)	0.84 (0.56-1.26)	0.87 (0.34-2.18)	

Adiponectin SNP T45G: 1 = T allele; 2 = G allele; $\chi^2 = 9.21$ for $\alpha = 0.01$ with df = 2; OR = odds ratio; NS = non significant; n = sample size.

5.4.9.7 Meta-analysis of adiponectin SNP T45G

The lack of association of the T45G alteration to disease risk in both cohorts presented is important as according to Hara *et al.* (2002) the variant allele has been associated with disease risk in the Japanese population (p value = 0.003) with an increased risk for carriers which are heterozygous (OR 1.41, 95% CI 1.06-1.88) as well as those homozygous for the variant allele (OR 1.70, 95% CI 1.09-2.65). The most striking finding upon comparison of the Southern African population to the Japanese population investigated by Hara *et al.* is the apparent rarity of the variant allele within the former. As presented in Table 5.25 the variant allele is present at a frequency of 0.01 and 0.02 in the patient and control cohorts respectively, whereas the Japanese population harbours this allele at frequencies of 0.35 within the diabetic cohort and 0.29 within the control cohort. Thus it may be hypothesised that this variant has an effect on disease susceptibility, however, as it is present at such a low frequency it is not discernible within the black Southern African cohort via the methods utilised in this investigation. However, the fact that the variant allele at this locus is present at a frequency of 0.13 in the Cuban diabetic cohort, and it is not associated with diabetic risk, indicates that it may not be a significant "at risk" allele within this population.

Another possible explanation for the absence of association of T2D risk to the G allele at T45G was originally posited in a population based investigation by Stumvoll *et al.* (2002). In this investigation the carriers of the variant allele in both the heterozygous and homozygous forms was associated with an increased BMI (p value = 0.02) when compared to carriers of the 1,1 genotype. Furthermore individuals harbouring the 2 allele were found to present with much lower levels of insulin sensitivity. Upon stratification of the population however, it was determined the effect of the variant allele was lost in individuals that had a prior family history of T2D. It was hypothesised by Stumvoll *et al.* that the susceptibility load that was inherited had a greater effect on disease risk than this one locus on its own.

As there were no homozygotes for the variant allele detected within the black Southern African cohort it was not possible to use the Wolfe's method to perform a meta-analysis. The Wolfe's method cannot be applied to a contingency table, as depicted in Equation 4.5 that contains cells with no data (Jackson *et al.*, 2004). This is evident upon investigation of Equation 4.7c where the variance is determined by the sum of the inverse sizes of the different cohorts investigated. If one of the cohorts were zero, calculation of the equation would result in the generation of a non-differentiable number.

Even upon the application of the Peto method of meta-analysis which is resistant to cells which do not harbour any information (Sweeting *et al.*, 2002), it was still not possible to utilise the cohorts investigated. Calculation of OR_i from a contingency table harbouring an empty cell results in the generation of either a zero or a non differentiable number. Whereas the latter cannot be utilised for obvious reasons, the former is also not informative as the natural logarithm of zero is non-differentiable. The cohorts could thus not be analysed via meta-analysis at the T45G locus by any comparable technique as utilised at the other loci.

5.4.10 G-11391A within the adiponectin gene

Although not an initial aim of this study, it was decided to screen the G-11391A locus within the Southern African cohorts in order to determine if it was significantly associated to disease risk. As discussed in Section 3.3.4 a specific haplotype generated utilising this alteration in conjunction with the C-11377G alteration has been associated with increased disease risk (Schwarz *et al.*, 2004) and it was reasoned that the role of this alteration in disease risk should be determined within the black Southern African cohorts investigated.

The G-11391A alteration within the adiponectin gene was screened via melting curve analysis following 35 cycles of amplification with a T_a of 62°C. Distinction of the peaks produced upon investigation of the negative differential graph for G-11391A was not as simple as that for the previous alterations as presented in Figure 5.9. This is possibly due to the alteration inducing only a slight change i.e. of 2-3°C in the melting curve of the 2,2 heterozygote. When detecting the 1,2 heterozygote the LC™ machine may not have been sensitive enough to differentiate the decrease in fluorescence of both the 1 allele and the 2 allele and thus produced only a single peak at the mean temperature of the two graphs. Individuals determined to be heterozygous were however screened via a sequencing

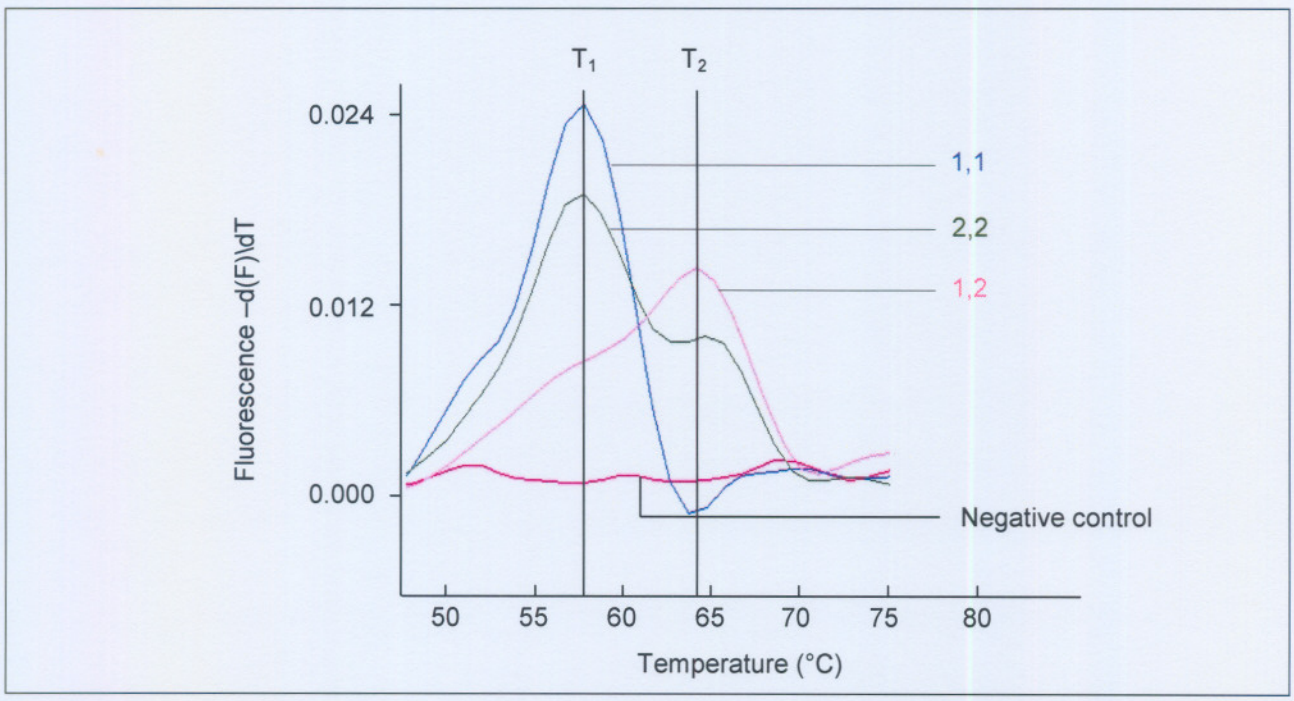
strategy[1] in order to validate this fact. Upon sequencing it was determined that these individuals were indeed heterozygous at the G-11391A locus.

As stated previously the sensor probe for this alteration is specific for the G allele. Hence the 1,1 genotype (pink curve in Figure 5.7) was assigned to the peak with the higher melting temperature (T_1 = 57.79°C, standard deviation = 0.63°C) and the 2,2 genotype to the peak with the lower temperature peak (T_2 = 64.36°C, standard deviation = 1.03°C) similar to the blue curve in Figure 5.7. In certain individuals it was possible to decipher two peaks at both melting temperatures (as represented by the green curve in Figure 5.7) for an individual heterozygous at this locus, and it was possible to calculate the following averages and standard deviations:

- heterozygote T_1 = 65.73°C, standard deviation = 0.48°C
- heterozygote T_2 = 57.93°C, standard deviation = 0.31°C

Due to the difficulty in peak differentiation as discussed in the previous paragraph the cohorts underwent screening for this alteration twice to ensure the correct assignment of genotypes. It will be useful for future investigations to elucidate the binding dynamics of this probe.

Figure 5.9: Diagrammatic representation of the differential graph of probe fluorescence versus temperature for G-11391A within the adiponectin gene



Adiponectin SNP G-11391A: 1 = G allele; 2 = A allele; 1,1 = homozygote for the 1 allele indicated by the red line; 1,2 = heterozygote indicated by the blue line; 2,2 = homozygote for 2 allele indicated by the green line; negative control indicated by purple line wherein DNA was replaced by deionised water; T_1 = temperature at which the 2,2 homozygote peaks; T_2 = temperature at which the 1,1 homozygote peaks.

[1] Sequencing strategy optimised and performed by U. Buro.

5.4.10.1 Adiponectin SNP G-11391A within the black Southern African diabetic cohort

The G-11391A alteration was investigated to determine the status of HW equilibrium at this locus. The population was in fact in HW equilibrium ($\chi^2 = 0.02$) as presented in Table 5.47 and therefore the assumptions associated with this hypothesis as delineated in Section 5.3.2 can be applied to the population under investigation.

Table 5.47: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African patient cohort for G-11391A

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	235	4	0	239
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.98	0.02	0.00	1.00
Expected numbers (E)	235.02	3.96	0.02	239.00
$\chi^2 = (O-E)^2/E$	0.00	0.00	0.02	0.02

Adiponectin SNP G-11391A: 1 = G allele; 2 = A allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.10.2 Adiponectin SNP G-11391A within the black Southern African control cohort

The cohort was in HW equilibrium ($\chi^2 = 0.30$) with regard to the G-11391A alteration as depicted in Table 5.48. Therefore determination of association of a specific genotype to disease risk can be undertaken in both the patient and control population as the genotype frequencies within the cohort are not currently affected by the factors discussed in Section 5.3.2.

Table 5.48: Chi-square test of goodness-of-fit to the HW proportions of the black Southern African control cohort for G-11391A

Genotypes	1,1	1,2	2,2	Total
Observed numbers (O)	215	16	0	231
Expected proportions	p^2	$2pq$	q^2	$p^2 + 2pq + q^2$
Frequency	0.93	0.07	0.00	1.00
Expected numbers (E)	215.28	15.45	0.28	231.00
$\chi^2 = (O-E)^2/E$	0.00	0.02	0.28	0.30

Adiponectin SNP G-11391A: 1 = G allele; 2 = A allele; χ^2 = Chi-square value; O = Observed numbers; E = Expected numbers; p = frequency of allele 1; q = frequency of allele 2.

5.4.10.3 Comparison of adiponectin SNP G-11391A between both black Southern African cohorts

The 1,1 homozygote presents with an OR > 1 as calculated in Table 5.49, and it is thus possible that this genotype is associated with greater risk, however the large range of the 95% CI means caution should be exercised with regard to this association. Via the utilisation of χ^2 analysis it was determined, that the heterozygotic genotype at G-11391A was significantly associated with protection towards T2D (p value < 0.01; OR = 0.22 (95% CI 0.07-0.69)) as indicated by the shaded cell within Table 5.49.

Table 5.49: Chi-square analysis for the comparison of the adiponectin genotype distribution at the G-11391A locus between the black Southern African patient and control cohorts

Cohort	1,1	1,2	2,2	p-value
Patient	0.98 (n=235)	0.02 (n=4)	0.00	<0.01
Control	0.93 (n=215)	0.07 (n=16)	0.00	
OR (95% CI)	4.37 (1.43-13.2)	0.22 (0.07-0.69)	Not calculated	

Adiponectin SNP G-11391A: 1 = G allele; 2 = A allele; 2,2 homozygous carriers of the variant allele at G-11391A and T45G were not detected upon analyses of the various cohorts and therefore the sum of homozygotes for the variant allele and heterozygotes were utilised for statistical analyses to prevent non-differentiation due to division by zero; $\chi^2 = 9.21$ for $\alpha = 0.01$ with df = 2; constitution of haplotype = C-11377G and G-11391A.; OR = odds ratio; n = sample size; the shaded cell indicates the evidence for association between the 1,2 genotype and a protective effect.

The fact that no homozygotes were determined within the cohorts analysed prevents the determination of whether the previous effect is stronger in individuals harbouring this genotype. However the relative rarity of this allele is antithetic, as generally if an alteration is associated with a protective factor it should be preferentially selected. T2D is generally a late onset disease, therefore usually not affecting the fecundity of the individual. It is thus reasonable to assume that this allele is in association with a protective effect against disease. As this allele is, however within a late onset disease susceptibility locus, the effect of possible adaptive forces is countered, thus preventing enrichment of this allele.

5.4.10.4 Comparison of adiponectin SNP G-11391A between the Cuban cohorts

As stated in Section 5.4.10 screening of the cohorts collected for SNP G-11391A was not an aim of this investigation. It was however studied within the black Southern African cohorts collected in this study due to the ancestral nature of this population as discussed in Section 5.1.1.1. It would therefore be possible to investigate the alteration in frequency of this SNP between the aforementioned population and the German population discussed by Schwarz *et al.* (2004). Financial and logistical constraints however, prevented the screening of the Cuban cohort for this alteration.

5.4.11 Haplotype frequencies at the adiponectin locus

It is possible for certain haplotypes within a gene to be in association with T2D even if the association of the individual genotypes are not significant (Horikawa *et al.*, 2000; Malecki *et al.* 2002; Cassell *et al.*, 2002). Investigation into the haplotype structure of the Southern African individuals was undertaken to determine if this was also true for the APM1 gene. The two promoter variants were utilised in the generation of haplotypes, as similar to UCSNP-44 no association was detected to the exonic variant. This strategy was undertaken to compare the results generated in this investigation to the findings presented by Schwarz *et al.* (2004).

5.4.11.1 Adiponectin haplotype analysis within the black Southern African cohorts

The 12 haplotype was significantly associated with a protective effect against T2D (p value < 0.01; OR = 0.22 (95% CI 0.06-0.78)) as depicted in the shaded row in Table 5.50. This strengthens the hypothesis that the variant allele at G-11391A is associated with a protective factor in the Southern African population.

Table 5.50: Chi-square analysis for the comparison of the adiponectin haplotype distribution between the black Southern African patient and control cohorts

Haplotypes	Patients Frequency (n)	Controls Frequency (n)	Chi square value	OR 95% CI
11	0.86 (398)	0.82 (375)	0.71	1.30 (0.91-1.85)
12	0.00 (3)	0.03 (13)	7.91	0.22 (0.06-0.78)
21	0.14 (63)	0.15 (68)	0.55	0.89 (0.61-1.29)
Total	*464*	*456*	*9.17*	---

χ^2 = 9.21 for α = 0.01 with df = 2; constitution of haplotype = C-11377G and G -11391A; n = individuals from whom haplotype data, reflecting genotypes at both loci, were generated; the shaded row indicates the evidence for association of the 12 haplotype with a protective effect.

5.4.11.2 Adiponectin haplotype combination analysis within the black Southern African cohorts

Based on the distribution of haplotype combinations as depicted in Table 5.51 the 11/12 haplotype combination (frequency = 0.06) was identified to represent a possible protective factor as highlighted in the shaded row in the aforementioned table. This notion is supported by the odds ratio value as well as its 95% CI which presents with an interval range sufficiently below one to indicate the aforementioned effect. However this alteration

is at a relatively low frequency within this cohort, indicating that it may not have a significant effect on a population level, but instead affords protection on the individual level. The identification of factors that act at the population level will require analysis of a larger cohort.

Table 5.51: Chi-square analysis for the comparison of the adiponectin haplotype combination distribution between the black Southern African patient and control cohorts

Haplotype Combination	Patients Frequency (n)	Controls Frequency (n)	Chi square value	OR 95% CI
11 / 11	0.72 (167)	0.67 (152)	0.98	1.28 (0.86-1.91)
11 / 12	0.01 (3)	0.06 (13)	7.91	0.21 (0.06-0.77)
11 / 21	0.26 (61)	0.25 (58)	0.07	1.04 (0.68-1.58)
21 / 21	0.01 (1)	0.02 (5)	3.28	0.19 (0.02-1.66)
Total	*232*	*228*	*12.24*	---

χ^2 = 11.34 for α = 0.01 with df = 3; constitution of haplotype = C-11377G and G-11391A; n = individuals for whom haplotype data were generated; the shaded block indicates the evidence for association of the 11/12 haplotype combination with a protective effect.

5.4.11.3 Comparison of adiponectin haplotypes and haplotype combinations to non-African populations

It was determined that the 11/12 haplotype (frequency = 0.06) represents association with a possible protective factor for T2D within the black South African population. This is not surprising as the variant allele at G-11391A by itself was determined to be associated with a similar effect. The fact that this association is maintained at the haplotype combination level only strengthens the protective role for this variant.

Comparison of the African data to that of the German population (Schwarz *et al.*, 2004) highlighted the fact that the 11/21 haplotype combination was linked to T2D susceptibility only in the European population. Due to the 11/21 haplotype combination being associated with disease risk in the German cohort while the 11/12 variant may be a protective factor in the black South African population, it could be hypothesised that a differential metabolic effect for genetic alterations in the APM1 gene with regard to T2D risk in these two populations exists. It is possible that the relative absence of variant alleles within the black South African diabetic cohort may be due to the fact that the 11/21 haplotype combination underwent a change in selective pressure upon migration into Europe, thus resulting in the disease risk with which it is currently associated.

This investigation highlights the population specific investigation of complex diseases such as T2D, which in turn impacts on the development of possible future management

strategies. A different role is therefore suggested for APM1 in T2D disease susceptibility within the black South African population. This alludes to the unfortunate fact that intervention strategies developed elsewhere may ultimately not be as effective in the Black South African population, as the genetic susceptibility factors may differ. The factors mentioned may however be effective to stratify personal T2D risk in preventative management.

5.5 GENOTYPE AND HAPLOTYPE RESULTS OF SOUTHERN AFRICAN AND CUBAN COHORTS AT SPECIFIC LOCI WITHIN THE CALPAIN 10 AND ADIPONECTIN GENES

Due to the current nature of T2D, this disease has been well studied within populations of developed countries as these individuals are at the highest risk of developing the disorder (King *et al.*, 1998). However this trend is rapidly changing and the necessity for investigating populations from the so-called developing countries is becoming apparent.

This investigation was undertaken to garner evidence that different populations have different risk profiles of disease susceptibility, even at loci that have been well documented within the developed countries. The following section presents an overview of the results in favour of this hypothesis.

5.4.1 Association to genetic variation within the calpain 10 gene

In Table 5.52 and 5.54 a summary is presented of the p- and OR values generated for the genotype frequency comparisons in both the black Southern African and Cuban cohorts at the loci discussed in Section 4.3. As previously discussed in Section 5.4.4.3 there is significant association between the 1,1 genotype at UCSNP-56 and a protective effect within the South African cohort.

Although there is significant variation within this cohort at the UCSNP-43 locus it was not possible to detect which genotype was responsible for this result. Alternatively the Cuban diabetic and control cohorts are not significantly different at the loci of UCSNP-43, -44 and -56. The UCSNP-63 locus presents with significant variation in the Cuban cohort however it is not possible to differentiate the origin of this effect due to the Cuban patient cohort not being in HW equilibrium as discussed in Section 5.4.5.6.

Table 5.52: Summary of genotype results generated at the various loci within the calpain 10 gene of both the black Southern African and Cuban cohorts

Southern African				Cuban			
SNP	*Gt*	*OR (95% CI)*	*p-value*	*SNP*	*Gt*	*OR (95% CI)*	*p-value*
UCSNP43	1,1	1.17 (0.77-1.78)	<0.001	UCSNP43	1,1	1.20 (0.85-1.68)	NS
	1,2	0.74 (0.48-1.15)			1,2	0.78 (0.55-1.11)	
	2,2	3.45 (0.71-16.8)			2,2	1.23 (0.56-2.68)	
UCSNP44	1,1	1.44 (0.80-2.59)	NS	UCSNP44	1,1	1.10 (0.74-1.63)	NS
	1,2	0.71 (0.39-1.29)			1,2	0.84 (0.56-1.27)	
	2,2	0.48 (0.04-5.43)			2,2	2.4 (0.48-11.9)	
UCSNP56	**1,1**	**0.25 (0.12-0.49)**	<0.0001	UCSNP56	1,1	0.94 (0.66-1.34)	NS
	1,2	0.86 (0.59-1.24)			1,2	0.84 (0.61-1.18)	
	2,2	**1.91 (1.32-2.77)**			2,2	1.37 (0.91-2.05)	
UCSNP63	1,1	0.82 (0.53-1.27)	NS	UCSNP63	1,1	0.76 (0.52-1.09)	**p<0.0001**
	1,2	0.99 (0.68-1.44)			1,2	1.04 (0.70-1.53)	
	2,2	1.18 (0.78-1.79)			2,2	3.59 (1.34-9.64)	

Odds ratio values highlighted in bold indicate the genotypes which have been significantly associated with modifying disease risk; Values within pink blocks (xxx) are associated with increased risk towards T2D whereas values within green blocks (yyy) are associated with protection against T2D; p-values highlighted in bold and within blue blocks (zzz) represent the p-value derived where one of the cohorts were not in HW equilibrium; Gt = genotype; NS = non significant; SNP = single nucleotide polymorphism; 95% CI = 95% confidence interval.

The OR and χ^2 values generated for the various haplotypes within the CAPN10 gene are outlined in Table 5.53. As previously discussed in Section 5.4.6.1, the 122 and 112 haplotypes present with the greatest association towards disease risk within the Southern African population while the Cubans did not present with any obvious associations. This may mean that these loci within CAPN10 are not important risk factors in T2D susceptibility. This fact further strengthens the discussion regarding variable disease susceptibility factors within various populations.

Table 5.53: Summary of haplotype results generated within the calpain 10 gene of both the black Southern African and Cuban cohorts

Southern African			Cuban		
Haplotypes	*χ^2 value*	*OR (95% CI)*	*Haplotypes*	*χ^2 value*	*OR (95% CI)*
121	1.05	1.13 (0.82-1.57)	121	0.04	0.97 (0.74-1.27)
122	**14.84**	**1.59 (1.22-2.08)**	**122**	**9.70**	**1.40 (1.01-1.94)**
111	3.99	0.73 (0.49-1.07)	111	0.93	0.90 (0.70-1.16)
112	**32.31**	**0.14 (0.06-0.31)**	112	5.74	3.12 (0.34-28.0)
221	0.30	0.72 (0.16-3.27)	221	0.07	0.90 (0.30-2.71)
211	1.59	0.81 (0.53-1.22)	211	1.28	0.87 (0.65-1.17)
222	ND	ND	222	1.29	ND
212	ND	ND	212	ND	0.97 (0.74-1.27)
Total	*54.08*	---	*Total*	*17.75*	---

Odds ratio values highlighted in bold indicate the haplotypes which have been significantly associated with modifying disease risk; Values within pink blocks (xxx) are associated with increased risk towards T2D whereas values within green blocks (yyy) are associated with protection against T2D; OR = odds ratio; 95% CI = 95% confidence interval; ND = non differentiable.

Upon meta-analysis of these various genotypes it became evident that although strong associations were detected within the various cohorts investigated, often these associations were absent upon meta-analytical analysis. This is significant to the investigation of susceptibility loci as ultimately important associations may be discarded as spurious upon meta-analytical analysis because it is not retained. If associations are maintained at this level it is important to determine the origin of this association. Population level effects however are just as important for the treatment of diabetes, as therapeutic strategies developed with regard to a specific population may not be useful in the treatment of T2D in another country or population. Therefore application of such strategies will ultimately not be useful to preventing disease risk. If the discussion presented in the introductory paragraph of this chapter is taken into consideration it is obvious that developing countries such as South Africa and Cuba cannot afford to implement such strategies.

5.4.2 Association to genetic variation within the adiponectin gene

At the APM1 locus, as presented within Table 5.54, it has been determined that the 1,2 genotype at the G-11391A locus is most likely to be associated with a protective factor towards T2D. The 1,1 genotype presents with increased risk according to the OR value described, however the range of the 95% CI is wide thus detracting from the strength of this association. It will be necessary in future to investigate a larger cohort in order to narrow this CI in order to definitively describe this association.

The role of this alteration is not discernible within the Cuban cohort, as it has not been screened at this locus. Comparison of the genotype frequencies at the C-11377G locus allowed for the calculation of a significant difference between the control and diabetic groups. At this SNP the 1,2 genotype was also associated with a protective factor whereas the 2,2 genotype was associated with increased risk according to the OR value > 1.

Table 5.54: Summary of genotype results generated at the various loci within the adiponectin gene of both the black Southern African and Cuban cohorts

Southern African				Cuban			
SNP	*Gt*	*OR*	*p-value*	*SNP*	*Gt*	*OR*	*p-value*
C-11377G	1,1	1.06 (0.70-1.59)	NS	C-11377G	1,1	1.24 (0.89-1.74)	<0.0001
	1,2	1.02 (0.68-1.54)			**1,2**	**0.68 (0.48-0.96)**	
	2,2	0.19 (0.02-1.67)			**2,2**	**2.54 (1.00-6.46)**	
T45G	1,1	1.55 (0.43-5.58)	NS	T45G	1,1	1.19 (0.81-1.75)	NS
	1,2	0.48 (0.11-1.94)			1,2	0.84 (0.56-1.26)	
	2,2	Not calculated			2,2	0.87 (0.34-2.18)	
G-11391A	**1,1**	**4.37 (1.43-13.2)**	<0.01	G-11391A	1,1	---	---
	1,2	**0.22 (0.07-0.69)**			1,2	---	
	2,2	Not calculated			2,2	---	

Odds ratio values highlighted in bold indicate the genotypes which have been significantly associated with disease risk; Values within pink blocks (xxx) are associated with increased risk towards T2D whereas values within green blocks (yyy) are associated with protection against T2D; Gt = genotype; NS = non significant.

As the genotypes present at G-11391A had not been screened within the Cuban cohort, it was not possible to present the haplotype data in a similar manner as Table 5.53, however due to the fact that there are differences in disease risk at the level of genotypes it is likely, as with the CAPN10 gene, that this is also the case at the haplotype level within the APM1 gene. As discussed in Section 5.4.11 the 12 haplotype and the 11/12 haplotype combination were associated with protection towards disease risk whereas the 11/21 haplotype combination was associated with increased risk towards T2D within the German population. The functional effect of these alterations will have to be detected in a future investigation.

Ultimately population level identification of susceptibility loci will decrease the overall cost of treating individuals affected by this disorder as certain therapeutic strategies can be ruled out. Further understanding of these effects could also result in the development of a genetic profile, which will place an individual at a greater risk. If this has been elucidated it is possible to diagnose an individual prior to the onset of disease. This in itself will decrease the disease load of the various health care services. The individual may be counselled to alter their lifestyle accordingly which ultimately may lead to the prevention of disease onset.

CHAPTER SIX

CONCLUSIONS

Energy metabolism is an integral biological process within all living organisms. It is therefore self-evident that many disorders e.g. T2D, which arise in humans are due to dysregulation in this metabolic process. This may be due to alterations in mRNA expression, protein expression, environmental factors or numerous other changes within the organism. In order to understand the methods by which this dysregulation affects cellular respiration, the basal metabolic state of unaffected individuals are compared to that of affected individuals. Via this process it is possible to identify important nodes in the disease pathogenesis network.

Current knowledge about the aetiology and pathogenesis of T2D has led to the description of a multifactorial model for disease susceptibility as presented in Figure 6.3 (see page 153). Barnett *et al.* (1981) and Bergman *et al.* (1981) have respectively described strong genetic and environmental components implicit in disease progression. This model is however broad and thus incomplete, as it has not yet been possible to define the specific interactions involved in T2D disease risk. Future investigation of the genetic aetiology of T2D will allow greater understanding of the role of genes, the environment and the interplay between these two entities.

Better comprehension of disease pathways will lead to the development of more efficient therapeutic strategies which, as discussed in Section 6.4, result in decreasing the overall financial burden of this disease (Rodgers *et al.*, 2004). Furthermore increased insight into the biological pathways involved in disease pathogenesis may result in reclassification of the diagnostic system for this group of disorders i.e. describing the disorder in terms of the pathway affected as opposed to the utilisation of symptoms such as hyperglycaemia (WHO Consortium, 1999).

To date numerous genes have been identified to be responsible for greater susceptibility to this disease (Froguel *et al.*, 1992; Horikawa *et al.*, 2000). Large levels of inter population variability have been determined at these loci (Fullerton *et al.*, 2002). This supports the

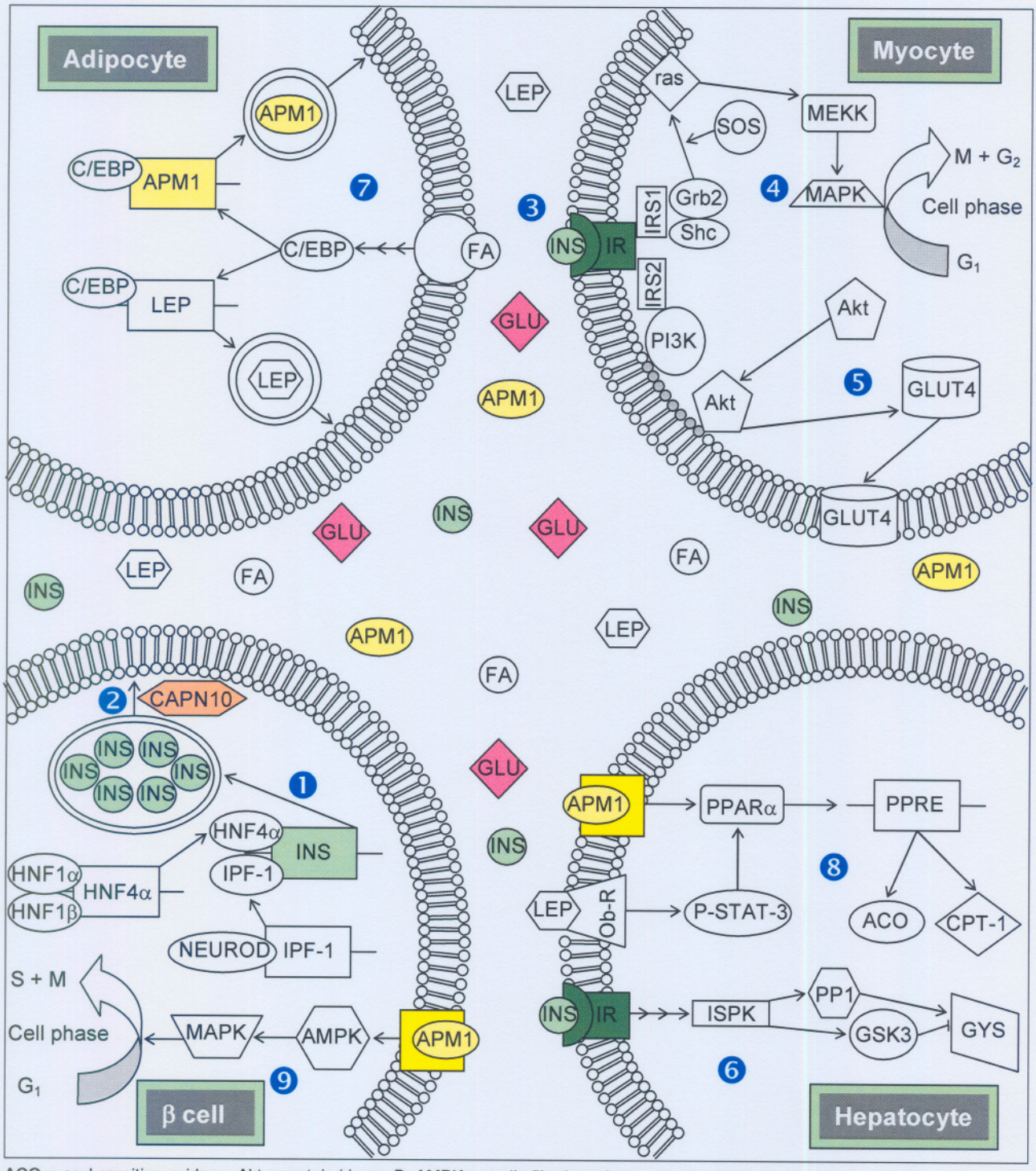
notion that the ethnic origin of an individual is integral to the correct assignment of genetic susceptibility towards the disease.

6.1 SIGNAL TRANSDUCTION AND ITS IMPORTANCE IN TYPE 2 DIABETES MELLITUS SUSCEPTIBILITY

The origin of this group of disorders are complex (WHO Consortium, 1999) as it is the cumulative effect of numerous interactions, which eventually cross a threshold level that result in the disorder. A model of the major molecular interactions occurring in a normoglycaemic individual is presented in Figure 6.1. A detailed description of each process in this figure is presented in Chapters Two and Three. It is important, however, to view this information in a systems biological framework given the aetiological complexity of this disorder. The major signalling processes under normoglycaemic conditions are presented in Figure 6.1 where references to specific pathways are indicated (❶ - ❾ in Figure 6.1) and discussed in the following paragraph. The same processes during hyperglycaemic conditions are presented in Figure 6.2 and are indicated (❶ - ❽ in Figure 6.2) but are discussed in the paragraph following Figure 6.1.

Under normoglycaemic conditions, the correct expression (❶) and secretion (❷) of insulin results in the appropriate activation of the insulin receptor (❸) in response to increased plasma glucose levels. Sensitivity towards insulin results in the transduction of signals which increase cellular growth (❹), glucose uptake (❺) and glycogen production (❻) as depicted within the myocyte and hepatocyte in Figure 6.1. Furthermore adipocytokines secreted by the adipocyte in response to fatty acids (❼) results in increased lipid breakdown (❽) as well as induction of cellular mitosis (❾) thus decreasing the plasma levels of both fatty acids and glucose. As depicted in Figure 6.1, it is no longer sufficient to merely concentrate on the so-called triumvirate of the hepatocyte, the β cell and the myocyte (DeFronzo, 1988) which are the major sites of these processes, when investigating T2D pathogenesis. At the very least the effects of molecules secreted by adipocytes should also be taken into consideration. Investigation of this cell type is by no means comprehensive and metabolomic strategies must be derived for future investigations into the mechanism of these disorders.

Figure 6.1: Model of the major intercellular signalling pathways under normoglycaemic conditions



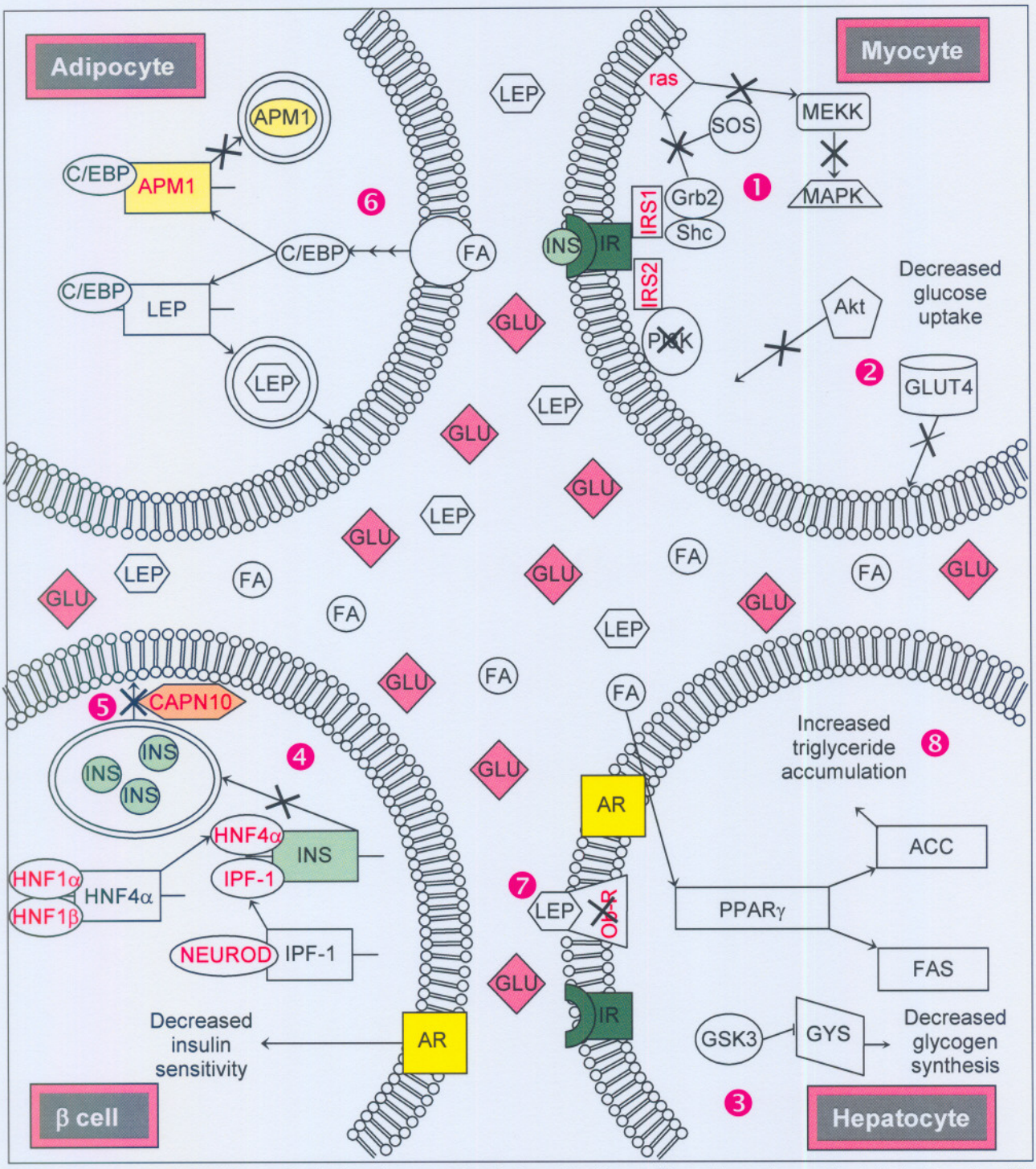
ACO = acyl carnitine oxidase; Akt = protein kinase B; AMPK = cyclic 5'-adenosine monophosphate protein kinase; APM1 = adiponectin indicated in yellow (□); CAPN10 = calpain 10 indicated in orange (□); C/EBP = CCAAT enhancer binding protein; Cell phases: G_1 = first growth phase; S = synthesis phase; G_2 = second growth phase; M = mitotic phase; CPT-1 = carnitine palmitoyl transferase 1; FA = fatty acid; GLU = glucose indicated in pink (□); GLUT 4 = glucose transporter 4; Grb-2 = growth factor receptor binding protein 2; GSK3 = glycogen synthase kinase 3; GYS = glycogen synthase; HNF1α = hepatocyte nuclear factor 1 alpha; HNF1β = hepatocyte nuclear factor 1 beta; HNF4α = hepatocyte nuclear factor 4 alpha; INS = insulin indicated in green (□); IPF-1 = insulin promoter factor 1; IR = insulin receptor; IRS1 = insulin receptor substrate 1; IRS2 = insulin receptor substrate 2; ISPK = insulin stimulated protein kinase; LEP = leptin; MAPK = mitogen activated protein kinase; MEKK = mitogen activated protein kinase kinase 1; NEUROD = neurogenin D; Ob-R = leptin receptor; PI3K = phosphatidylinositol 3'-kinase; PP1 = protein phosphatase 1; PPARα = peroxisome proliferator activated receptor alpha; PPRE = peroxisome proliferator activated receptor response element; P-STAT-3 = phosphorylated signal transducers and activators of transcription 3; ras = cellular form of the rat sarcoma proto-oncogene; Shc = *src* homology 2 domain-containing oncogenic protein; SOS = son of sevenless.

Whilst undergoing hyperglycaemic conditions, as discussed in Section 2.1.4.2, insulin resistance arises from defects in the signalling pathways of insulin itself (❶), which prevents the correct uptake (❷), storage (❸) and metabolism of glucose within the myocyte (Reavan, 1988). Although not indicated in the hepatocyte illustrated in Figure 6.2 a similar loss of insulin-induced effects occurs as in the myocyte (❶). In association with this loss of sensitivity there is no increase in glycogen synthesis (❸), thus maintaining gluconeogenesis and increasing plasma glucose levels (Dent *et al.*, 1990). Insulin resistance is however, not the only pathway of disease pathogenesis. Insufficient insulin secretion due to decreased expression (❹) of this hormone or dysregulation of the secretory process (❺) will result in insufficient plasma levels of insulin thus affecting the appropriate hyperinsulinaemic response towards hyperglycaemia. Incorrect expression (❻), or resistance to adipocytokines (❼), result in increased free fatty acid plasma levels as well as accumulation of triglycerides (❽) in non-adipocytes causing apoptosis (Unger *et al.*, 1999).

A proposed model outlining the progression of disease pathogenesis is presented in Figure 6.3, wherein it is apparent that genetic susceptibility may have effects at numerous levels such as decreased glycogen synthesis (Dent *et al.*, 1990), increased lipolysis (Thomas *et al.*, 1979), peripheral insulin resistance (DeFronzo *et al.*, 1985), β cell loss (Unger *et al.*, 1999) as well as β cell compensation (Chiu *et al.*, 2000). Both genetic (Barnett *et al.,* 1981) and environmental factors (Eriksson and Lindgärde, 1991) can have an effect on the various processes discussed which ultimately result in T2D.

Insulin resistance is generally the central cause of disease pathogenesis and can arise due to the direct effects of genes (Vestergaard *et al.*, 1993) or indirectly due to increased genetic susceptibility to obesity which itself induces insulin resistance (Dohm *et al.*, 1988). Barring these effects the resistance may also arise due to ageing. Generally in insulin resistant individuals the β cells are able to compensate by producing greater levels of insulin (Bergman *et al.*, 1981). Genetic susceptibility to β cell loss can, however, prevent this compensatory mechanism from being activated (Hegele *et al.*, 1999).

Figure 6.2: Model of the major intercellular signalling pathways under hyperglycaemic conditions

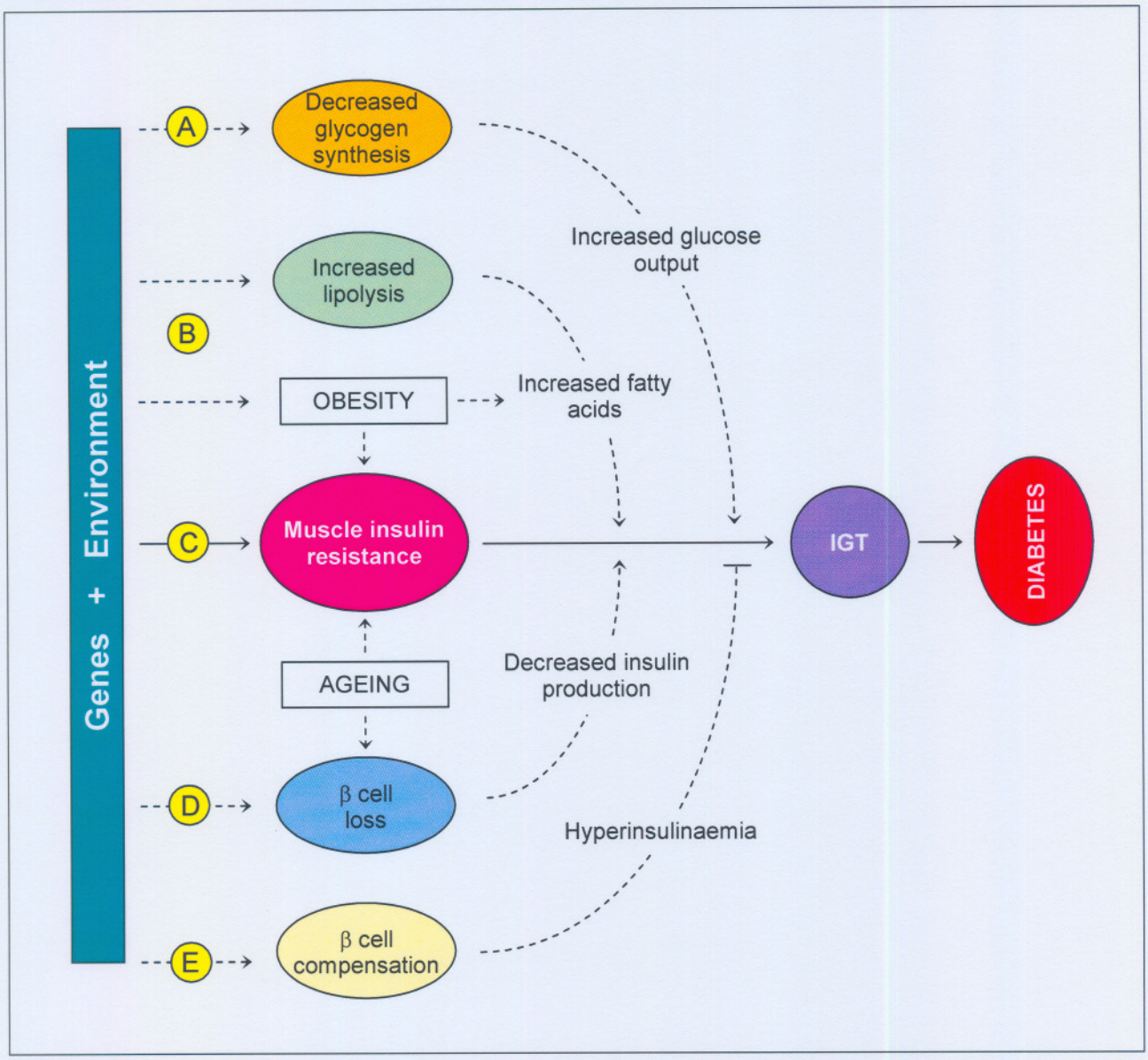


Akt = protein kinase B; AMPK = cyclic 5'-adenosine monophosphate protein kinase; AR = adiponectin receptor indicated in yellow ( ); CAPN10 = calpain 10 indicated in orange ( ); C/EBP = CCAAT enhancer binding protein; FA = fatty acid; GLU = glucose indicated in pink ( ); GLUT 4 = glucose transporter 4; Grb-2 = growth factor receptor binding protein 2; GSK3 = glycogen synthase kinase 3; GYS = glycogen synthase; HNF1α = hepatocyte nuclear factor 1 alpha; HNF1β = hepatocyte nuclear factor 1 beta; HNF4α = hepatocyte nuclear factor 4 alpha; INS = insulin indicated in green ( ); IPF-1 = insulin promoter factor 1; IRS1 = insulin receptor substrate 1; IRS2 = insulin receptor substrate 2; LEP = leptin; MEKK = mitogen activated protein kinase kinase 1; NEUROD = neurogenin D; Ob-R = leptin receptor; PPARγ = peroxisome proliferator activated receptor gamma; PI3K = phosphatidylinositol 3'-kinase; ras = cellular form of the rat sarcoma proto-oncogene; Shc = *src* homology 2 domain-containing oncogenic protein; SOS = son of sevenless; text in red (xxx) indicates protein products of the most important genes which were discussed in Chapter Three with reference to disease risk.

Although the initial lesion may be peripheral insulin resistance, this effect is further aggravated by increased fatty acid levels within the circulatory system, which may arise from obesity due to an increased uptake of these molecules from the digestive system

(Randle *et al.*, 1963) or genetic alterations in fatty acid transporters such as the FABP2 gene (Baier *et al.*, 1995). Another source of these molecules is the concomitant increase in lipolysis due to the dysregulation of insulin (Thomas *et al.*, 1979).

Figure 6.3: Proposed model of the development of type 2 diabetes mellitus



IGT = impaired glucose tolerance; Letters in yellow circles represent the possible genetic loci involved in each pathway of disease progression: (A) includes the genes encoding protein phosphatase 1 regulatory subunit 2 (PPP1R2), protein phosphatase 1 regulatory subunit 3A (PPP1R3A), glycogen synthase (GYS1), glycogen synthase kinase 3 alpha (GSK3α); (B) includes the genes encoding fatty acid binding protein 2 (FABP2), leptin receptor (LEPR), adiponectin (APM1), tumour necrosis factor α (TNFα), interleukin-6 (IL6); (C) includes the genes encoding insulin receptor substrate 1 (IRS-1), insulin receptor substrate 2 (IRS-2), mitogen activated protein kinase 8-interacting protein 1 (MAPK8IP1), ras associated with diabetes (RAD), adiponectin (APM1), tumour necrosis factor α (TNFα), interleukin-6 (IL6); (D) includes the genes encoding hepatocyte nuclear factor 4 alpha (HNF4α), hepatocyte nuclear factor 1 alpha (HNF1α), hepatocyte nuclear factor 1 beta (HNF1β), insulin promoter factor (IPF-1), neurogenin D1 (NEUROD1), leptin receptor (LEPR); (E) includes the genes encoding glucokinase and calpain 10 (CAPN10).

Fatty acids are able to induce expression of proteins such as ACC and FAS, which causes accumulation of triglycerides within myocytes, hepatocytes and the islets of Langerhans (Unger *et al.*, 1999). Upon reaching a threshold these cells undergo apoptosis which in turn decreases insulin production. Genetic defects in the MODY genes discussed in

Table 5.52: Summary of genotype results generated at the various loci within the calpain 10 gene of both the black Southern African and Cuban cohorts

Southern African				Cuban			
SNP	***Gt***	***OR (95% CI)***	***p-value***	***SNP***	***Gt***	***OR (95% CI)***	***p-value***
UCSNP43	1,1	1.17 (0.77-1.78)	<0.001	UCSNP43	1,1	1.20 (0.85-1.68)	NS
	1,2	0.74 (0.48-1.15)			1,2	0.78 (0.55-1.11)	
	2,2	3.45 (0.71-16.8)			2,2	1.23 (0.56-2.68)	
UCSNP44	1,1	1.44 (0.80-2.59)	NS	UCSNP44	1,1	1.10 (0.74-1.63)	NS
	1,2	0.71 (0.39-1.29)			1,2	0.84 (0.56-1.27)	
	2,2	0.48 (0.04-5.43)			2,2	2.4 (0.48-11.9)	
UCSNP56	**1,1**	**0.25 (0.12-0.49)**	<0.0001	UCSNP56	1,1	0.94 (0.66-1.34)	NS
	1,2	0.86 (0.59-1.24)			1,2	0.84 (0.61-1.18)	
					2,2	1.37 (0.91-2.05)	
UCSNP63	1,1	0.82 (0.53-1.27)	NS	UCSNP63	1,1	0.76 (0.52-1.09)	**p<0.0001**
	1,2	0.99 (0.68-1.44)			1,2	1.04 (0.70-1.53)	
	2,2	1.18 (0.78-1.79)			2,2	3.59 (1.34-9.64)	

Odds ratio values highlighted in bold indicate the genotypes which have been significantly associated with modifying disease risk; Values within pink blocks ( ) are associated with increased risk towards T2D whereas values within green blocks ( ) are associated with protection against T2D; p-values highlighted in bold and within blue blocks ( ) represent the p-value derived where one of the cohorts were not in HW equilibrium; Gt = genotype; NS = non significant; SNP = single nucleotide polymorphism; 95% CI = 95% confidence interval.

The OR and χ^2 values generated for the various haplotypes within the CAPN10 gene are outlined in Table 5.53. As previously discussed in Section 5.4.6.1, the 122 and 112 haplotypes present with the greatest association towards disease risk within the Southern African population while the Cubans did not present with any obvious associations. This may mean that these loci within CAPN10 are not important risk factors in T2D susceptibility. This fact further strengthens the discussion regarding variable disease susceptibility factors within various populations.

Table 5.53: Summary of haplotype results generated within the calpain 10 gene of both the black Southern African and Cuban cohorts

Southern African			Cuban		
Haplotypes	***χ^2 value***	***OR (95% CI)***	***Haplotypes***	***χ^2 value***	***OR (95% CI)***
121	1.05	1.13 (0.82-1.57)	121	0.04	0.97 (0.74-1.27)
111	3.99	0.73 (0.49-1.07)	111	0.93	0.90 (0.70-1.16)
112	**32.31**	**0.14 (0.06-0.31)**	112	5.74	3.12 (0.34-28.0)
221	0.30	0.72 (0.16-3.27)	221	0.07	0.90 (0.30-2.71)
211	1.59	0.81 (0.53-1.22)	211	1.28	0.87 (0.65-1.17)
222	ND	ND	222	1.29	ND
212	ND	ND	212	ND	0.97 (0.74-1.27)
Total	*54.08*	---	*Total*	*17.75*	---

Odds ratio values highlighted in bold indicate the haplotypes which have been significantly associated with modifying disease risk; Values within pink blocks ( ) are associated with increased risk towards T2D whereas values within green blocks ( ) are associated with protection against T2D; OR = odds ratio; 95% CI = 95% confidence interval; ND = non differentiable.

Table 5.54: Summary of genotype results generated at the various loci within the adiponectin gene of both the black Southern African and Cuban cohorts

Southern African				Cuban			
SNP	***Gt***	***OR***	***p-value***	***SNP***	***Gt***	***OR***	***p-value***
C-11377G	1,1	1.06 (0.70-1.59)	NS	C-11377G	1,1	1.24 (0.89-1.74)	<0.0001
	1,2	1.02 (0.68-1.54)			**1,2**	**0.68 (0.48-0.96)**	
	2,2	0.19 (0.02-1.67)					
T45G	1,1	1.55 (0.43-5.58)	NS	T45G	1,1	1.19 (0.81-1.75)	NS
	1,2	0.48 (0.11-1.94)			1,2	0.84 (0.56-1.26)	
	2,2	Not calculated			2,2	0.87 (0.34-2.18)	
G-11391A			<0.01	G-11391A	1,1	---	---
	1,2	**0.22 (0.07-0.69)**			1,2	---	
	2,2	Not calculated			2,2	---	

Odds ratio values highlighted in bold indicate the genotypes which have been significantly associated with disease risk; Values within pink blocks ( ) are associated with increased risk towards T2D whereas values within green blocks ( ) are associated with protection against T2D; Gt = genotype; NS = non significant.

As the genotypes present at G-11391A had not been screened within the Cuban cohort, it was not possible to present the haplotype data in a similar manner as Table 5.53, however due to the fact that there are differences in disease risk at the level of genotypes it is likely, as with the CAPN10 gene, that this is also the case at the haplotype level within the APM1 gene. As discussed in Section 5.4.11 the 12 haplotype and the 11/12 haplotype combination were associated with protection towards disease risk whereas the 11/21 haplotype combination was associated with increased risk towards T2D within the German population. The functional effect of these alterations will have to be detected in a future investigation.

Ultimately population level identification of susceptibility loci will decrease the overall cost of treating individuals affected by this disorder as certain therapeutic strategies can be ruled out. Further understanding of these effects could also result in the development of a genetic profile, which will place an individual at a greater risk. If this has been elucidated it is possible to diagnose an individual prior to the onset of disease. This in itself will decrease the disease load of the various health care services. The individual may be counselled to alter their lifestyle accordingly which ultimately may lead to the prevention of disease onset.

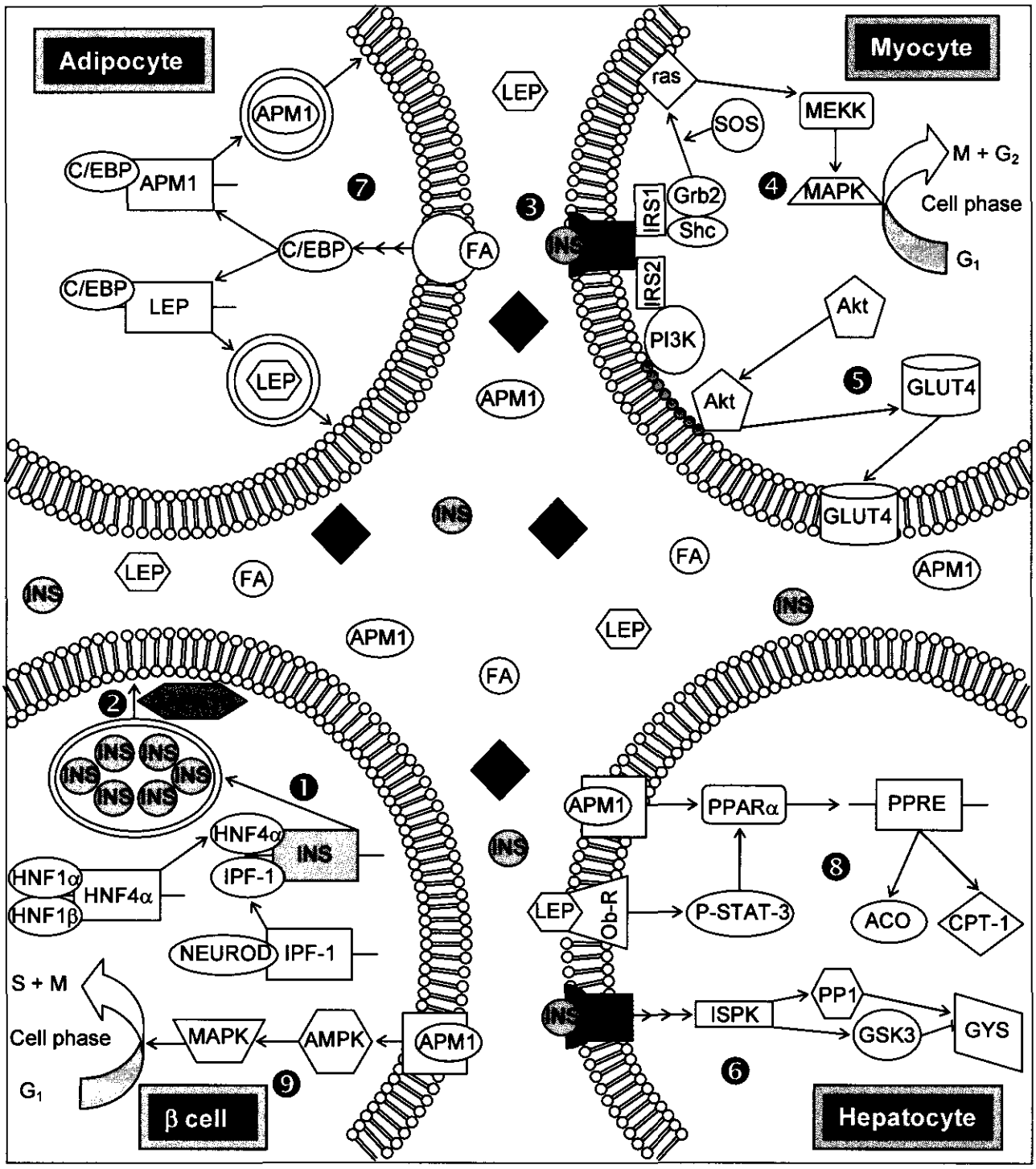
notion that the ethnic origin of an individual is integral to the correct assignment of genetic susceptibility towards the disease.

6.1 SIGNAL TRANSDUCTION AND ITS IMPORTANCE IN TYPE 2 DIABETES MELLITUS SUSCEPTIBILITY

The origin of this group of disorders are complex (WHO Consortium, 1999) as it is the cumulative effect of numerous interactions, which eventually cross a threshold level that result in the disorder. A model of the major molecular interactions occurring in a normoglycaemic individual is presented in Figure 6.1. A detailed description of each process in this figure is presented in Chapters Two and Three. It is important, however, to view this information in a systems biological framework given the aetiological complexity of this disorder. The major signalling processes under normoglycaemic conditions are presented in Figure 6.1 where references to specific pathways are indicated (❶ - ❾ in Figure 6.1) and discussed in the following paragraph. The same processes during hyperglycaemic conditions are presented in Figure 6.2 and are indicated (❶ - ❽ in Figure 6.2) but are discussed in the paragraph following Figure 6.1.

Under normoglycaemic conditions, the correct expression (❶) and secretion (❷) of insulin results in the appropriate activation of the insulin receptor (❸) in response to increased plasma glucose levels. Sensitivity towards insulin results in the transduction of signals which increase cellular growth (❹), glucose uptake (❺) and glycogen production (❻) as depicted within the myocyte and hepatocyte in Figure 6.1. Furthermore adipocytokines secreted by the adipocyte in response to fatty acids (❼) results in increased lipid breakdown (❽) as well as induction of cellular mitosis (❾) thus decreasing the plasma levels of both fatty acids and glucose. As depicted in Figure 6.1, it is no longer sufficient to merely concentrate on the so-called triumvirate of the hepatocyte, the β cell and the myocyte (DeFronzo, 1988) which are the major sites of these processes, when investigating T2D pathogenesis. At the very least the effects of molecules secreted by adipocytes should also be taken into consideration. Investigation of this cell type is by no means comprehensive and metabolomic strategies must be derived for future investigations into the mechanism of these disorders.

Figure 6.1: Model of the major intercellular signalling pathways under normoglycaemic conditions



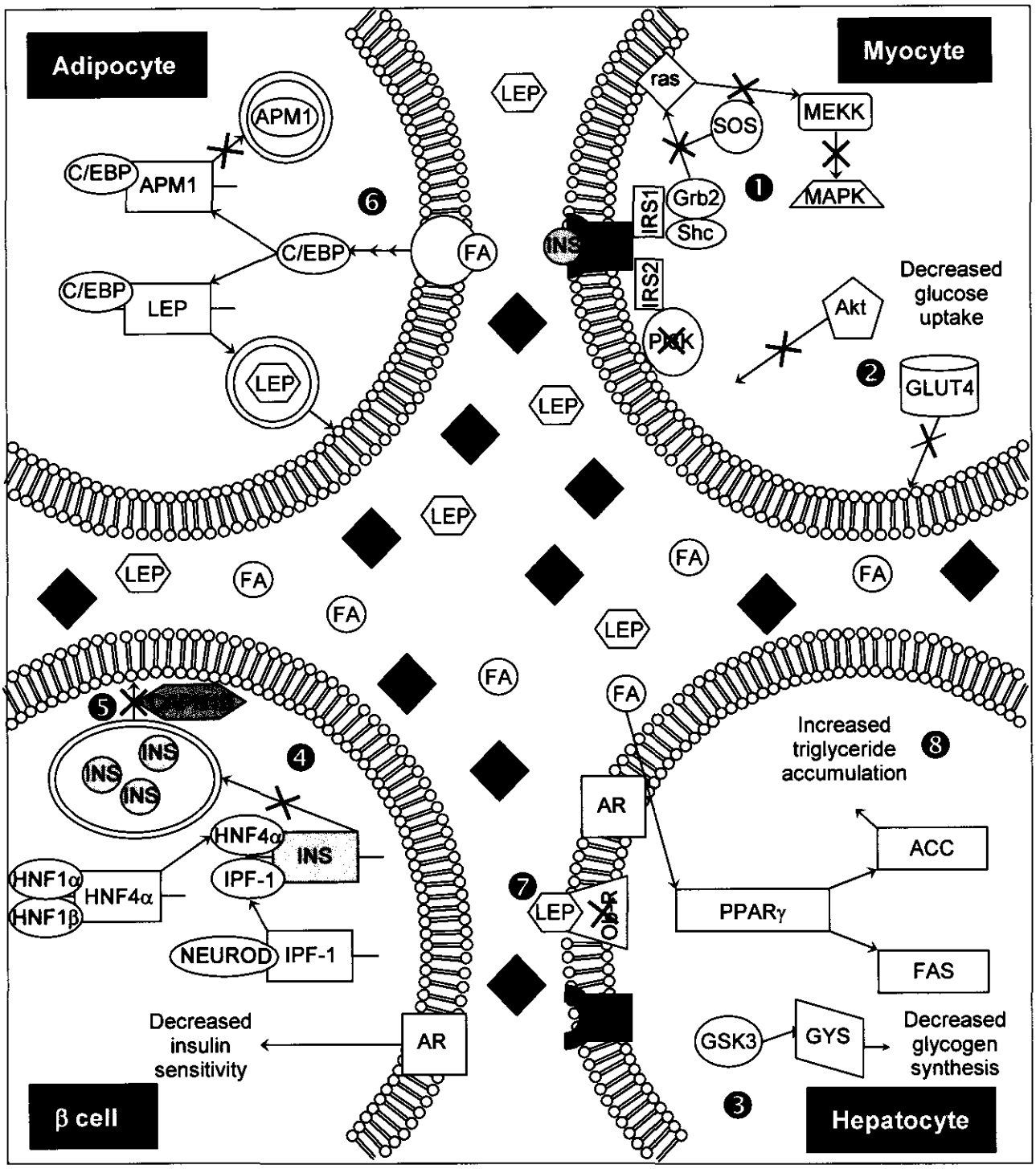
ACO = acyl carnitine oxidase; Akt = protein kinase B; AMPK = cyclic 5'-adenosine monophosphate protein kinase; APM1 = adiponectin indicated in yellow (□); CAPN10 = calpain 10 indicated in orange (■); C/EBP = CCAAT enhancer binding protein; Cell phases: G_1 = first growth phase; S = synthesis phase; G_2 = second growth phase; M = mitotic phase; CPT-1 = carnitine palmitoyl transferase 1; FA = fatty acid; GLU = glucose indicated in pink (■); GLUT 4 = glucose transporter 4; Grb-2 = growth factor receptor binding protein 2; GSK3 = glycogen synthase kinase 3; GYS = glycogen synthase; HNF1α = hepatocyte nuclear factor 1 alpha; HNF1β = hepatocyte nuclear factor 1 beta; HNF4α = hepatocyte nuclear factor 4 alpha; INS = insulin indicated in green (■); IPF-1 = insulin promoter factor 1; IR = insulin receptor; IRS1 = insulin receptor substrate 1; IRS2 = insulin receptor substrate 2; ISPK = insulin stimulated protein kinase; LEP = leptin; MAPK = mitogen activated protein kinase; MEKK = mitogen activated protein kinase kinase 1; NEUROD = neurogenin D; Ob-R = leptin receptor; PI3K = phosphatidylinositol 3'-kinase; PP1 = protein phosphatase 1; PPARα = peroxisome proliferator activated receptor alpha; PPRE = peroxisome proliferator activated receptor response element; P-STAT-3 = phosphorylated signal transducers and activators of transcription 3; ras = cellular form of the rat sarcoma proto-oncogene; Shc = *src* homology 2 domain-containing oncogenic protein; SOS = son of sevenless.

Whilst undergoing hyperglycaemic conditions, as discussed in Section 2.1.4.2, insulin resistance arises from defects in the signalling pathways of insulin itself (❶), which prevents the correct uptake (❷), storage (❸) and metabolism of glucose within the myocyte (Reavan, 1988). Although not indicated in the hepatocyte illustrated in Figure 6.2 a similar loss of insulin-induced effects occurs as in the myocyte (❶). In association with this loss of sensitivity there is no increase in glycogen synthesis (❸), thus maintaining gluconeogenesis and increasing plasma glucose levels (Dent *et al.*, 1990). Insulin resistance is however, not the only pathway of disease pathogenesis. Insufficient insulin secretion due to decreased expression (❹) of this hormone or dysregulation of the secretory process (❺) will result in insufficient plasma levels of insulin thus affecting the appropriate hyperinsulinaemic response towards hyperglycaemia. Incorrect expression (❻), or resistance to adipocytokines (❼), result in increased free fatty acid plasma levels as well as accumulation of triglycerides (❽) in non-adipocytes causing apoptosis (Unger *et al.*, 1999).

A proposed model outlining the progression of disease pathogenesis is presented in Figure 6.3, wherein it is apparent that genetic susceptibility may have effects at numerous levels such as decreased glycogen synthesis (Dent *et al.*, 1990), increased lipolysis (Thomas *et al.*, 1979), peripheral insulin resistance (DeFronzo *et al.*, 1985), β cell loss (Unger *et al.*, 1999) as well as β cell compensation (Chiu *et al.*, 2000). Both genetic (Barnett *et al.*, 1981) and environmental factors (Eriksson and Lindgärde, 1991) can have an effect on the various processes discussed which ultimately result in T2D.

Insulin resistance is generally the central cause of disease pathogenesis and can arise due to the direct effects of genes (Vestergaard *et al.*, 1993) or indirectly due to increased genetic susceptibility to obesity which itself induces insulin resistance (Dohm *et al.*, 1988). Barring these effects the resistance may also arise due to ageing. Generally in insulin resistant individuals the β cells are able to compensate by producing greater levels of insulin (Bergman *et al.*, 1981). Genetic susceptibility to β cell loss can, however, prevent this compensatory mechanism from being activated (Hegele *et al.*, 1999).

Figure 6.2: Model of the major intercellular signalling pathways under hyperglycaemic conditions

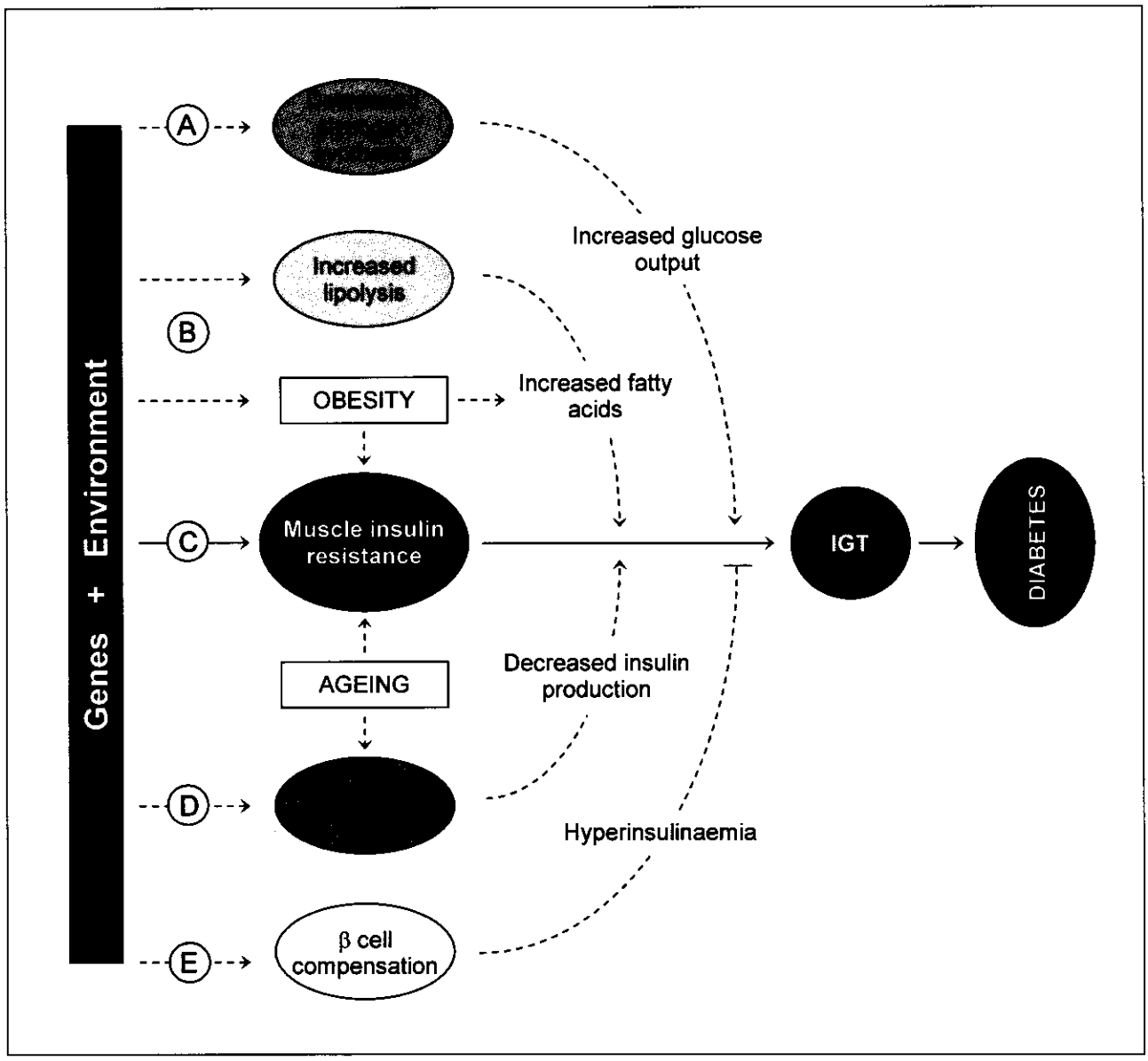


Akt = protein kinase B; AMPK = cyclic 5'-adenosine monophosphate protein kinase; AR = adiponectin receptor indicated in yellow (□) CAPN10 = calpain 10 indicated in orange (■); C/EBP = CCAAT enhancer binding protein; FA = fatty acid; GLU = glucose indicated in pink (■); GLUT 4 = glucose transporter 4; Grb-2 = growth factor receptor binding protein 2; GSK3 = glycogen synthase kinase 3; GYS = glycogen synthase; HNF1α = hepatocyte nuclear factor 1 alpha; HNF1β = hepatocyte nuclear factor 1 beta; HNF4α = hepatocyte nuclear factor 4 alpha; INS = insulin indicated in green (□); IPF-1 = insulin promoter factor 1; IRS1 = insulin receptor substrate 1; IRS2 = insulin receptor substrate 2; LEP = leptin; MEKK = mitogen activated protein kinase kinase 1; NEUROD = neurogenin D; Ob-R = leptin receptor; PPARγ = peroxisome proliferator activated receptor gamma; PI3K = phosphatidylinositol 3'-kinase; ras = cellular form of the rat sarcoma proto-oncogene; Shc = *src* homology 2 domain-containing oncogenic protein; SOS = son of sevenless; text in red (xxx) indicates protein products of the most important genes which were discussed in Chapter Three with reference to disease risk.

Although the initial lesion may be peripheral insulin resistance, this effect is further aggravated by increased fatty acid levels within the circulatory system, which may arise from obesity due to an increased uptake of these molecules from the digestive system

(Randle *et al.*, 1963) or genetic alterations in fatty acid transporters such as the FABP2 gene (Baier *et al.*, 1995). Another source of these molecules is the concomitant increase in lipolysis due to the dysregulation of insulin (Thomas *et al.*, 1979).

Figure 6.3: Proposed model of the development of type 2 diabetes mellitus



IGT = impaired glucose tolerance; Letters in yellow circles represent the possible genetic loci involved in each pathway of disease progression: Ⓐ includes the genes encoding protein phosphatase 1 regulatory subunit 2 (PPP1R2), protein phosphatase 1 regulatory subunit 3A (PPP1R3A), glycogen synthase (GYS1), glycogen synthase kinase 3 alpha (GSK3α); Ⓑ includes the genes encoding fatty acid binding protein 2 (FABP2), leptin receptor (LEPR), adiponectin (APM1), tumour necrosis factor α (TNFα), interleukin-6 (IL6); Ⓒ includes the genes encoding insulin receptor substrate 1 (IRS-1), insulin receptor substrate 2 (IRS-2), mitogen activated protein kinase 8-interacting protein 1 (MAPK8IP1), ras associated with diabetes (RAD), adiponectin (APM1), tumour necrosis factor α (TNFα), interleukin-6 (IL6); Ⓓ includes the genes encoding hepatocyte nuclear factor 4 alpha (HNF4α), hepatocyte nuclear factor 1 alpha (HNF1α), hepatocyte nuclear factor 1 beta (HNF1β), insulin promoter factor (IPF-1), neurogenin D1 (NEUROD1), leptin receptor (LEPR); Ⓔ includes the genes encoding glucokinase and calpain 10 (CAPN10).

Fatty acids are able to induce expression of proteins such as ACC and FAS, which causes accumulation of triglycerides within myocytes, hepatocytes and the islets of Langerhans (Unger *et al.*, 1999). Upon reaching a threshold these cells undergo apoptosis which in turn decreases insulin production. Genetic defects in the MODY genes discussed in

Section 2.1.4.3.1 can also directly affect insulin secretion, which refers to a major lesion resulting in T2D and MODY dependent on the potency of the mutation (Hegele *et al.*, 1999). Long term hyperinsulinaemia may also result in the degradation of β cells (Kahn *et al.*, 1993). These factors therefore prevent the appropriate hyperinsulinaemic response to hyperglycaemia thus causing increased acceleration of disease progression.

Decreased glycogen synthesis may arise due to mutations within the regulatory processes of GYS1 or alternatively due to a genetic or environmental impairment in signal transduction (Vaag *et al.*, 1992). The resultant increased glucose output further aggravates the hyperglycaemia caused by the decreased sensitivity towards insulin. Whereas a hyperinsulinaemic response may be able to overcome this effect, the decreased levels of this hormone, due to β cell loss, prevents this from occurring. Ultimately this results in IGT which if left untreated progresses to T2D.

Due to the fact that the major defect within this group of disorders is that of errant signal transduction, it is becoming increasingly apparent that in future investigations it will be necessary to implement system biological strategies in order to elucidate the mechanisms of disease action. Elucidation of the exact genetic sequence of an organism is no longer sufficient. The intrinsic regulation of these genetic elements is becoming more important for understanding biological systems (Pennisi, 2004). The implementation of metabolomic and signalomic tools in the study of T2D is thus integral for investigation of disease risk.

This investigation in itself, however creates a plethora of difficulties as at present the necessary standards and controls do not exist for these types of analyses. Earlier 'omic' strategies such as microarray experiments were inherently non-comparable as manufacturing processes of commercial chips as well as initial sample collection were not standardised (Tan *et al.*, 2003). Development of a reliable metabolomic strategy will firstly require elucidation of the possible loci of internal and extraneous variability. The effects of these processes will in turn have to be countered in order to increase the signal to noise ratio of the various tests.

One of the major sources of metabolomic variability is genetic adaptation, which is inherent to the ancestry of an individual (Mishmar *et al.*, 2003). The evidence generated in this investigation alludes to the fact that differential risk patterns are present within different populations at specific T2D susceptibility loci.

6.2 EVIDENCE GENERATED FROM ANALYSIS OF THE CALPAIN 10 GENE

As previously discussed the exact role of CAPN10 in disease pathogenesis has not yet been fully elucidated, however a recent investigation with regard to mechanisms of insulin secretion has suggested a possible pathway. A protein termed islet cell autoantigen (ICA) 512, through an interaction with β-syntrophin-utrophin complexes, is hypothesised to anchor secretory granules to the actin cytoskeleton in neuroendocrine cells such as β cells. Glucose stimulation of β cells results in the calcium dependent cleavage of ICA512 via μ-calpain (also termed calpain 2) thus allowing release of the secretory granules and the concomitant secretion of insulin (Ort *et al.*, 2001). A possible positive feedback mechanism was observed upon stimulation via insulin as β-syntrophin is dephosphorylated via a PP1 mediated pathway which allows for more efficient cleavage of ICA512. It is thus plausible that if there is some lesion within the calpain there will be decreased secretion of insulin resulting in hyperglycaemia. CAPN10 may be involved within this process, either as a regulatory enzyme of calpain 2 or may itself cleave ICA512 in a non-calcium dependent manner. The role of CAPN10 within this process must be defined in the future (Ort *et al.*, 2001).

The association of a protective effect with the wild type allele at UCSNP-56 within the black South African cohorts investigated, as presented in Section 5.4.4.3, was surprising although this strengthens the hypothesis that the mechanism of disease onset is different among separate populations. Harbouring the variant allele at this locus in the black Southern African population, does not indicate increased risk but rather it increases the threshold of disease occurrence. If these loci are ever used as markers of disease risk they will only be applicable to non-African populations such as the Mexican American and European populations. For example a black Southern African individual may harbour the European "at risk" allele at UCSNP-43 as described by Horikawa *et al.* (2000) and may be at no greater risk of developing T2D. If the individual furthermore has the wild type allele at UCSNP-56 it may be protected against disease risk and thus the application of a preventative strategy developed in a different population would be unnecessary and costly.

In the Cuban groups studied however the association to UCSNP-56 was absent whereas significant variation was detected at the UCSNP-63 locus. The reason for this difference is not obvious due to the absence of HW equilibrium within the Cuban diabetic cohort. As

discussed in Section 5.4.5.6 this may be due to association between this SNP and disease risk, however further investigation is required before this may be stated definitively.

Both sets of these results would have been obscured if the Cuban and black Southern African cohorts had been combined. This was evident upon meta-analysis as presented in Sections 5.4.4.7 and 5.4.5.7. Inclusion of individuals in the investigation of T2D susceptibility loci, with no regard for ethnicity, may result in certain important risk factors being overlooked. This finding underscores the need for comparison of ethnically matched cohorts in studying T2D susceptibility loci.

As discussed in Section 5.4.6.1 and 5.4.6.3 similar haplotypes were detected in the black Southern African and Cuban cohorts as had been observed in the Mexican American, German and Finnish populations (Horikawa *et al.*, 2000). A significant deviation to this trend was, however, the association of the 112 haplotype to a protective effect within the black Southern African population investigated. This association was unexpected as the haplotype in the black Southern African control cohort was present at a frequency of 0.10 and not at a frequency less than 0.05 as in the ethnic groups investigated by Horikawa *et al.* (2000). Previously reported "rare" haplotypes are thus present at significant levels within the black Southern African population as compared to the populations investigated by Horikawa *et al.* (2000). This may be due to the differential disease risk patterns which have altered haplotype frequencies between different populations over time. In future a larger sample of individuals should be investigated to verify the frequency of this haplotype within the extended population. It will also be useful to screen the most ancient lineages within macrohaplogroup L to determine the haplotype frequencies within these populations. This information may be utilised to investigate if any variation exists in the haplotype frequencies between these ancestral populations and other modern lineages. If this is the case, the driving force behind allele frequency variation must be determined in order to elucidate important mechanisms of disease evolution.

The argument for the utilisation of ethnically matched cohorts in T2D susceptibility analysis is significantly strengthened upon the discovery of the absence of association to certain "at-risk" haplotype combinations in the black South African and Cuban cohorts investigated with T2D. The protective effect associated with the 112/112 haplotype combination as well as the association of disease risk with the 122/122 haplotype combination in the black Southern African cohort, as discussed in Section 5.4.6.2, indicates the strength of the allele at UCSNP-56 as this is the only variant between these two combinations. The

preliminary association of the 122/122 haplotype combination with disease risk in the Cuban cohort, as indicated in Section 5.4.6.4, garners greater evidence for the possible role of this haplotype in disease risk. This overlap in "at risk" haplotype combinations between these two populations may be due to the fact that the African population is one of the founder populations of modern Cubans (Torroni *et al.*, 1995).

There are numerous reasons for this discovery of variable susceptibility however the most likely explanation is that different genetic structures within CAPN10 are responsible for disease risk in the black Southern African and Cuban populations. This would be expected as natural selection cannot act as strongly on susceptibility loci as it does on loci which have direct positive or negative effects, thus resulting in more variation within these aforementioned loci during human evolution as well as preventing fixation of any single allele (Fullerton *et al.*, 2002).

6.3 EVIDENCE GENERATED FROM ANALYSIS OF THE ADIPONECTIN GENE

Investigation into certain SNPs within the APM1 gene revealed that these alterations also presented with population specific effects. Within the Cuban cohorts investigated as discussed in Section 5.4.8.6, the 2,2 genotype at the C-11377G locus was associated with increased risk. A protective effect was, however, associated with individuals harbouring the heterozygote at this position. This alludes to the C-11377G locus being a significant point of reference for disease risk in the Cuban population. Meta-analysis of this locus furthermore indicated that the 1 allele was possibly associated with a protective effect in the three groups investigated. The wild type homozygote has the maximum protective effect compared to the mutant homozygote however the heterozygote has a similar effect as the former. This indicates that it is not a dosage dependent effect. Functional analysis of this alteration in the future may result in the elucidation of its role in disease risk.

The 1,2 genotype at the G-11391A locus was marginally associated with a protective effect within the Southern African cohort investigated as discussed in Section 5.4.10.3. It is not clear whether this effect is due to the variant allele specifically or the heterozygote genotype itself as no mutant homozygotes were detected in the black Southern African cohort investigated. The association is marginal and due to the low proportion of individuals that do not harbour the wild type homozygote genotype, it is unlikely that this is a significant candidate for population level assessment of risk. It may however, possibly be utilised as an indicator of individual disease risk. It is necessary that this effect be further

investigated in the Cuban population, which may harbour greater levels of the variant allele, as this will assist in determining its role in disease pathogenesis.

The absence of any association at the T45G locus in any of the groups studied in this investigation is also relevant due to the strong association present in the Japanese population (Hara *et al.*, 2002). This may be indicative of the fact that the Asian specific risk factors are different from those present in the Southern African and Cuban populations, thus increasing the evidence that specific populations have specific patterns of genetic susceptibility.

In the course of the investigation by Schwarz *et al.*, (2004) it was determined that certain haplotypes within the promoter region of APM1 were associated with either higher or lower levels of adiponectin within the German population. Individuals harbouring the 'low adiponectin' haplotypes were thus affected by long-term hypoadiponectinaemia. It was also determined that these individuals were the most likely to progress to the next glucose impairment class. If the mechanism of the 'low adiponectin' haplotypes previously discussed is taken into consideration the association with disease progression becomes self-evident. Lower levels of adiponectin, which were associated with certain APM1 promoter haplotypes, result in less activation of the various signalling pathways. The loss of the increase in glucose utilisation thus causes resistance toward the signal induced by insulin, resulting in the various peripheral cells becoming resistant toward this hormone. Furthermore the accumulation of triglycerides within the β cells results in apoptosis of these cells, thus overall cell mass is decreased (Unger *et al.*, 1999). This in turn results in the loss of correct insulin production. The diabetic phenotype is thus concomitantly due to an increase in insulin resistance as well as a decrease in insulin production.

Haplotype analysis as discussed in Section 5.4.11 was only undertaken within the black Southern African cohort, as discussed in Section 5.4.10. It was determined that the 12 haplotype was significantly associated with protection towards T2D as presented in Section 5.4.11.1. This association may be due to the mechanism discussed in the previous paragraph however functional analysis should be undertaken in the future in order to determine the phenotypic effect of harbouring this haplotype.

The 11/21 haplotype combination associated with disease risk within the German population (Schwarz *et al.*, 2004) was present within the black Southern African cohort but was not significantly associated with disease risk. Alternatively the 11/12 haplotype

combination was associated with protection towards disease risk. This indicates that there may be differential effects in disease pathogenesis with regard to this locus between the two populations. Future functional analysis of the promoter region of this gene will assist in unravelling the origins of this association.

It may be that various other genetic lesions within the gene itself, which are not in linkage disequilibrium with the alterations investigated, are involved. However the major aim of this investigation was to uncover the variation present at the well-described loci, within this gene in the black Southern African population. Future elucidation of novel alterations associated with disease risk will strengthen the hypothesis of a novel genetic aetiology for disease risk in a specific population thus indicating that differential preventative measures are required in treating T2D. It is possible that a global investigation of disease susceptibility loci can prevent the detection of population specific effects which are important indicators of disease risk (Liu *et al.*, 2004). As the origins of complex disorders are inherently complicated it is necessary to identify as many contributing factors as possible in order to elucidate a true effect.

6.4 IMPORTANCE OF GENETIC ANCESTRY IN SUSCEPTIBILITY TO TYPE 2 DIABETES MELLITUS

The fact that the genetic background of an individual affects the pathogenic expression of certain alterations has recently been underscored in an investigation by Brown *et al.* in 2002. His investigation was in reference to the mitochondrial genetic background however similar systems must be present in the nuclear genome.

Certain polymorphic alterations within the mitochondrial genome have become fixed over time in the various human populations. It is therefore possible to utilise these alterations to determine the specific lineage to which an individual belongs. Due to adaptation of the mitochondrial genome within the human population, only certain alterations are today prevalent within specific geographical regions. The haplogroup of an individual is therefore determined by screening its mitochondrial genome for these specific alterations (Johnson *et al.*, 1983). Brown *et al.* in 2002 detected a possible predisposing effect that a haplogroup may have on a mutation. It was determined that the specific alterations termed 11778A and 14484C, which both result in Leber's hereditary optic neuropathy (LHON) occurred 3-6 and 9 times more often in haplogroup J than in other haplogroups, respectively. Different genetic risk factors have alternative phenotypes depending on the genetic background upon which expression takes place.

Although previously believed to be polymorphic it is now becoming apparent as stated in the previous paragraph, that the various alterations which define specific haplogroups, are in fact undergoing adaptation. Via comparison of individuals from different geographic locations Mishmar *et al.* in 2003 were able to determine that individuals from the arctic regions present with increased uncoupling of mitochondrial function. The adaptive advantage becomes obvious when it is taken into consideration that greater heat generation via proton uncoupling is associated with this process. Alternatively in the African populations, mitochondrial coupling is more efficient, as there is very little selective advantage for greater heat generation in the more temperate environment in which these populations evolved. If adaptation is present at this level the various nuances of cellular metabolism must provide a veritable gamut of opportunities for metabolic control.

Different populations thus have altered basal metabolic rates due to evolutionary adaptation to various environmental factors. If so much diversity is present within the smaller mitochondrial genome the addition of the nuclear genome must add an almost impossible level of complexity as evidenced by the discussion in Section 6.2 and 6.3. Moreover, the black African population belong to macrohaplogroup L which harbours the most ancient human lineage as well as having the highest level of variation (Chen *et al.*, 1995) it becomes nearly impossible to envision all the possible factors involved in disease susceptibility.

Alternatively the origins of the Cuban population are genetically heterogeneous (Torroni *et al.*, 1995). The genomic background of this population is therefore more complex due to admixture effects. The group investigated, although not currently affected by the effects of migration as determined by the presence of HW equilibrium at all loci investigated except that of UCSNP-63, is derived from the original founding populations discussed in Section 5.1.1.2 (Torroni *et al.*, 1995), making it an important and worthwhile population to study.

It was thus vital to investigate population specific cohorts to remove the extraneous effects of inter population variability, in order to define genetic as well as environmental risk factors. This allowed for the baseline comparison of the individuals and patients to ensure that any alteration detected in the metabolism of an individual was not due to population specific variation but rather to actual association with the disease.

In doing so it was possible to elucidate that there is a differential pattern in the disease risk imparted by specific reported SNPs in the CAPN10 and APM1 genes between the cohorts investigated in the study presented here. A possible explanation for this is that the various

epistatic as well as epigenetic effects, which have arisen from the widely different environments in which these populations have evolved, may have altered the mechanisms of disease expression. As discussed in Section 6.1 aberrant signal transduction is central to T2D susceptibility. There are thus numerous points at which such factors could affect disease risk. Elucidation of these factors will however require the implementation of system biological approaches in order to generate an unbiased view of disease pathogenesis.

The results presented do however allude to the unfortunate fact that intervention strategies developed elsewhere may ultimately not be as effective in these populations, given that different population specific risk factors are at play, as discussed with reference to HIV/AIDS by Gonzalez *et al.* (2001). These genetic factors may, however, be effective to stratify personal T2D risk in preventative management. Alternatively stated it may be possible to specifically develop personalised treatment that can be rapidly optimised. This will decrease health care costs as it will not be necessary to begin treatment utilising general treatment strategies which may not be relevant to a specific population.

It becomes necessary, therefore to investigate T2D susceptibility within a specific population. Ultimately any strategy incorporated to treat this disorder will require a deep understanding of disease susceptibility within the population. Most so-called Third World countries cannot afford to implement First World strategies, which may have little relevance to the treatment of T2D within these developing lands. It is therefore important to elucidate the population specific genetic variables involved in order to develop significant methods of stemming the flow of this burgeoning epidemic.

CHAPTER SEVEN

REFERENCES

7.1 GENERAL REFERENCES

Abbott C., Piaggio G., Ammendola R., Solomon E., Povey S., Gounari F., De Simone V. and Cortese R. Mapping of the gene TCF2 coding for the transcription factor LFB3 to human chromosome 17 by polymerase chain reaction. *Genomics*, **8**, 165-167, 1990.

Alcolado J.C. and Alcolado R. Importance of maternal history of non-insulin dependant diabetic patients. *BMJ.*, **302**, 1178-1180, 1991.

Almind K., Bjørbæk C., Vestergaard H., Hansen T., Echwald S. and Pedersen O. Amino acid polymorphisms of insulin receptor substrate-1 in non-insulin-dependent diabetes mellitus. *Lancet*, **342**, 828-832, 1993.

Almind K., Inoue G., Pedersen O. and Kahn C.R. A common amino acid polymorphism in insulin receptor substrate-1 causes impaired insulin signaling: evidence from transfection studies. *J. Clin. Invest.*, **97**, 2569-2575, 1996.

Araki E., Lipes M.A., Patti M.-E., Brüning J.C., Haag B., Johnson R.S. and Kahn C.R. Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. *Nature*, **372**, 186-190, 1994.

Argyrokastritis A., Kamakari S., Kapsetaki M., Kritis A., Talianidis I. and Moschonas N.K. Human hepatocyte nuclear factor-4 (hHNF-4) gene maps to 20q12-q13.1 between PLCG1 and D20S17. *Hum. Genet.*, **99**, 233-236, 1997.

Arribas M., Valverde A.M., Burks D., Klein J., Farese R.V., White M.F. and Benito M. Essential role of protein kinase Cζ in the impairment of insulin-induced glucose transport in IRS-2 deficient brown adipocytes. *FEBS. Lett.*, **536**, 161-166, 2003.

Bach I., Mattei M.-G., Cereghini S. and Yaniv M. Two members of an HNF1 homeoprotein family are expressed in human liver. *Nucleic Acids Res.*, **19**, 3553-3559, 1991.

Backer J.M., Myers M.G., Sun X.-J., Chin D.J., Shoelson S.E., Miralpeix M. and White M.F. Association of IRS-1 with the insulin receptor and the phosphatidylinositol 3'-kinase: Formation of binary and ternary signalling complexes in intact cells. *J. Biol. Chem.*, **268**, 8204-8212, 1993.

Baier L.J., Permana P.A., Yang X., Pratley R.E., Hanson R.L., Shen G-Q., Mott D., Knowler W.C., Cox N.J., Horikawa Y., Oda N., Bell G.I. and Bogardus C.A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. *J. Clin. Invest.*, **106**, R69-R73, 2000.

Baier L.J., Sacchettini J.C., Knowler W.C., Eads J., Paolisso G., Tataranni P.A., Mochizuki H., Bennett P.H., Bogardus C. and Prochazka M. An amino acid substitution in the human intestinal fatty acid binding protein is associated with increased fatty acid binding, increased fat oxidation and insulin resistance. *J. Clin. Invest.*, **95**, 1281-1287, 1995.

Ballinger S.W., Shoffner J.M., Hedaya E.V., Trounce I., Polak M.A., Koontz D.A. and Wallace D.C. Maternally transmitted diabetes and deafness associated with a 10.4 kb mitochondrial DNA deletion. *Nat. Genet.*, **1**, 11-15, 1992.

Baltensperger K., Kozma L.M., Cherniack A.D., Klarlund J.K., Chawla A., Banerjee U. and Czech M.P. Binding of the Ras activator son of sevenless to insulin receptor substrate-1 signaling complexes. *Science*, **260**, 1950-1952, 1993.

Bandyopadhyay G., Standaert M.L., Zhao L., Yu B., Avignon A., Galloway L., Karnam P., Moscat J. and Farese R.V. Activation of protein kinase C (α, β, and ζ) by insulin in 3T3/L1 cells: Transfection studies suggest a role for PKC-ζ in glucose transport. *J. Biol. Chem.*, **272**, 2551-2558, 1997.

Barker D.J.P., Bull A.R., Osmond C. and Simmonds S.J. Fetal and placental size and risk of hypertension in adult life. *BMJ.*, **301**, 259-262, 1990.

Barnett A.H., Eff C., Leslie R.D.G. and Pyke D.A. Diabetes in identical twins: A study of 200 pairs. *Diabetologia*, **20**, 87-93, 1981.

Baroni M.G., D' Andrea M.P., Montali A., Pannitteri G., Barilla F., Campagna F., Mazzei E., Lovari S., Seccareccia F., Campa P.P., Ricci G., Pozzilli P., Urbinati G. and Arca M. A common mutation of the insulin receptor substrate-1 gene is a risk factor for coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.*, **19**, 2975-2980, 1999.

Bastard J.-P., Maachi M., van Nhieu J.T., Jardel C., Bruckert E., Grimaldi A., Robert J.-J., Capeau J. and Hainque B. Adipose tissue IL-6 content correlates with resistance to insulin activation of glucose uptake both *in vivo* and *in vitro*. *J. Clin. Endocrinol. Metab.*, **87**, 2084-2089, 2002.

Bates S.H., Stearns W.H., Dundon T.A., Schubert M., Tso A.W.K., Wang Y., Banks A.S., Lavery H.J., Haq A.K., Maratos-Flier E., Neel B.G., Schwartz M.W. and Myers M.G. STAT3 signalling is required for leptin regulation of energy balance but not reproduction. *Nature,* **421**, 856-859, 2003.

Bektas A., Suprenant M.E., Wogan L.T., Plengvidhya N., Rich S.S., Warram J.H., Krolewski A.S. and Doria A. Evidence of a novel type 2 diabetes locus 50 cM centromeric to NIDDM2 on chromosome 12q. *Diabetes*, **48**, 2246-2251, 1999.

Bell G.I., Xiang K., Newman M.V., Wu S., Wright L.G., Fajans S.S., Spielman R.S. and Cox N.J. Gene for non-insulin-dependant diabetes mellitus (maturity-onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 20q. *Proc. Natl. Acad. Sci. USA.*, **88**, 1484-1488, 1991.

Benatar S.R. Health care reform and the crisis of HIV and AIDS in South Africa. *N. Engl. J. Med.*, **351**, 81-92, 2004.

Bergman R.N., Phillips L.S. and Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and β-cell glucose sensitivity from the response to intravenous glucose. *J. Clin. Invest.*, **68**, 1456-1467, 1981.

Bland J.M. and Altman D.G. The odds ratio. *BMJ.*, **320**, 1468, 2000.

Bland J.M. and Altman D.G. Multiple significance tests: the Bonferroni method. *BMJ.*, **310**, 170, 1995.

Bottazzo G.F. β–cell damage in diabetic insulinitis: are we approaching a solution? *Diabetologia*, **26**, 241-249, 1984.

Bowcock A.M., Kidd J.R., Mountain J.L., Hebert J.M., Carotenuto L., Kidd K.K. and Cavalli-Sfroza L.L. Drift, admixture and selection in human evolution: a study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA.*, **88**, 839-843, 1991.

Brown M.D., Starikovskaya E., Derbeneva O., Hosseini S., Allen J.C., Mikhailovskaya I.E., Sukernik R.I. and Wallace D.C. The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup. *Hum. Genet.*, **110**, 130-138, 2002.

Brown P.M., Tompkins C.V., Juul S. and Sönsken P.H. Mechanism of action of insulin in diabetic patients: a dose related effect on glucose production and utilisation. *BMJ.*, **1**, 1239-1242, 1978.

Brozinick J.T. and Birnbaum M.J. Insulin, but not contraction, activates Akt/PKB in isolated rat skeletal muscle. *J. Biol. Chem.*, **273**, 14679-14682, 1998.

Burks D.J., de Mora J.F., Schubert M., Withers D.J., Myers M.G., Towery H.H., Altamuro S.L., Flint C.L. and White M.F. IRS-2 pathways integrate female reproduction and energy homeostasis. *Nature*, **407**, 377-382, 2000.

Cann R.L., Stoneking M. and Wilson A.C. Mitochondrial DNA and human evolution. *Nature*, **325**, 31-36, 1987.

Carey A.L., Lamont B., Andrikopoulus S., Koukoulas I., Proietto J. and Febbraio M.A. Interleukin-6 gene expression is increased in insulin-resistant rat skeletal muscle following insulin stimulation. *Biochem. Biophys. Res. Commun.*, **302**, 837-840, 2003.

Cassell P.G., Jackson A.E., North B.V., Evans J.C., Syndercombe-Court D., Phillips C., Ramachandran A., Snehalatha C., Gelding S.V., Vijayaravaghan S., Curtis D. and Hitman G.A. Haplotype combinations of calpain 10 gene polymorphisms associate with increased risk of impaired glucose tolerance and type 2 diabetes in South Indians. *Diabetes*, **51**, 1622-1628, 2002.

Chen Y.H., Hansen L., Chen M.X., Bjørbæk C., Vestergaard H., Hansen T., Cohen P.T.W. and Pedersen O. Sequence of the human glycogen-associated regulatory subunit of type 1 protein phosphatase and analysis of its coding region and mRNA level in muscle from patients with NIDDM. *Diabetes*, **43**, 1234-1241, 1994.

Chen Y-S., Torroni A., Excoffier L., Santachiara-Benerecetti A.S. and Wallace D.C. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.*, **57**, 133-149, 1995.

Chiu K.C., Chuang L.-M., Yoon C. and Saad M.F. Hepatic glucokinase promoter polymorphism is associated with hepatic insulin resistance in Asian Indians. *BMC Genet.*, **1**, 2-7, 2000.

Chrichton N. Odds ratio. *J. Clin. Nursing.*, **10**, 257-269, 2001.

Christianson T.W. and Clayton D.A. A tridecamer DNA sequence supports human mitochondrial RNA 3'-end formation *in vitro*. *Mol. Cell. Biol.*, **8**, 4502-4509, 1988.

Chung W.K., Power-Kehoe L., Chua M., Chu F., Aronne L., Huma Z., Sothern M., Udall J.N., Kahle B. and Leibel R.L. Exonic and intronic sequence variation in the human leptin receptor gene (LEPR). *Diabetes*, **46**, 1509-1511, 1997.

Clausen J.O., Hansen T., Bjørbæk C., Echwald S.M., Urhammer S.A., Rasmussen S., Andersen C.B., Hansen L., Almind K., Winther K., Haraldsdottir J., Borch-Johnsen K. and Pedersen O. Insulin resistance: interactions between obesity and a common variant of insulin receptor substrate-1. *Lancet,* **346**, 397-402, 1995.

Cohen B., Novick D. and Rubinstein M. Modulation of insulin activities by leptin. *Science*, **274**, 1185-1188, 1996.

Congia M., Patel S., Cope A.P., De Virgiliis S. and Sonderstrup G. T cell epitopes of insulin defined in HLA-DR4 transgenic mice are derived from preproinsulin and proinsulin. *Proc. Natl. Acad. Sci. USA.*, **95**, 3833-3838, 1998.

Couce M.E., Burguera B., Parisi J.E., Jensen M.D. and Lloyd R.V. Localization of leptin receptor in the human brain. *Neuroendocrinology*, **66**, 145-150, 1997.

Cox N.J. Challenges in identifying genetic variation affecting susceptibility to type 2 diabetes: examples from studies of the calpain-10 gene. *Hum. Mol. Genet.*, **10**, 2301-2305, 2001.

Croall D.E. and Demartino G.N. Calcium-activated neutral protease (calpain) system: structure, function and regulation. *Physiol. Rev.*, **71**, 813-847, 1991.

DeFronzo R.A., Jacot E., Jequier E., Maeder E., Wahren J. and Felber J.P. The effect of insulin on the disposal of intravenous glucose. *Diabetes*, **30**, 1000-1007, 1981.

DeFronzo R.A., Gunnarsson R., Björkman O., Olsson M. and Wahren J. Effects of insulin on peripheral and sphlanic glucose metabolism in noninsulin-dependant (Type II) diabetes mellitus. *J. Clin. Invest.*, **76**, 149-155, 1985.

DeFronzo R.A. The triumvirate: β-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes*, **37**, 667-687, 1988.

DeFronzo R.A., Ferrannini E. and Simonson D.C. Fasting hyperglycemia in non-insulin-dependant diabetes mellitus: Contributions of excessive hepatic glucose production and impaired tissue glucose uptake. *Metabolism*, **38**, 387-395, 1989.

Degos L., Chaventré A. and Jacquard A. Migration, selection and histocompatibility. *J. Hum. Evol.*, **8**, 795-797, 1979.

Dent P., Lavoinne A., Nakielny S., Caudwell F.B., Watt P. and Cohen P. The molecular mechanism by which insulin stimulates glycogen synthesis in mammalian skeletal muscle. *Nature*, **348**, 302-308, 1990.

DerSimonian R. and Laird N. Meta-analysis in clinical trials. *Control Clin. Trials.*, **7**, 177-188, 1986.

De Silva S.N.T., Weerasuriya N., De Alwis N.M.W., De Silva M.W.A. and Fernando D.J.S. Excess maternal transmission and familial aggregation of type 2 diabetes in Sri Lanka. *Diabetes Res. Clin. Prac.*, **58**, 173-177, 2002.

Dohm G.L., Tapscott E.B., Pories W.J., Dabbs D.J., Flickinger E.G., Meelheim D., Fushiki T., Atkinson S.M., Elton C.W. and Caro J.F. An *in vitro* human muscle preparation suitable for metabolic studies. *J. Clin. Invest.*, **82**, 486-494, 1988.

Doria A., Caldwell J.S., Ji L., Reynet C., Rich S.S., Weremowicz S., Morton C.C., Warram J.H., Kahn C.R. and Krolewski A.S. Trinucleotide repeats at the *rad* locus: Allele distributions in NIDDM and mapping to a 3-cM region on chromosome 16q. *Diabetes*, **44**, 243-247, 1995.

Ducimetiere P., Eschwege E., Papoz L., Richard J.L., Claude J.R. and Rosselin G. Relationship of plasma insulin levels to the incidence of myocardial infarction and coronary heart disease mortality in a middle aged population. *Diabetologia*, **19**, 205-210, 1980.

Ebina Y., Ellis L., Jarnagin K., Edery M., Graf L., Clauser E., Ou J., Masiarz F., Kan Y.W., Goldfine I.D., Roth R.A. and Rutter W.J. The human insulin receptor cDNA: The structural basis for hormone-activated transmembrane signaling. *Cell*, **40**, 747-758, 1985.

Eck M.J., Dhe-Paganon S., Trüb T., Nolte R.T. and Shoelson S.E. Structure of the IRS-1 PTB domain bound to the juxtamembrane region of the insulin receptor. *Cell*, **85**, 695-705, 1996.

Ehrmann D.A., Schwarz P.E.H., Hara M., Tang X., Horikawa Y., Imperial J., Bell G.I. and Cox N.J. Relationship of calpain-10 genotype to phenotypic features of polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **87**, 1669-1673, 2002.

Elbein S.C., Chu W., Ren Q., Hemphill C., Schay J., Cox N.J., Hanis C.L. and Hasstedt S.J. Role of Calpain-10 gene variants in familial type 2 diabetes in Caucasians. *J. Clin. Endocrinol. Metab.*, **87**, 650-654, 2002.

Embi N., Rylatt D.B. and Cohen P. Glycogen synthase kinase-3 from rabbit skeletal muscle: Separation from cyclic-AMP-dependent protein kinase and phosphorylase kinase. *Eur. J. Biochem.*, **107**, 519-527, 1980.

Erasmus R.T., Blanco Blanco E., Okesina A.B., Mesa A.J., Gqweta Z. and Matsha T. Importance of family history in type 2 black South African diabetic patients. *Postgrad. Med. J.*, **77**, 323-325, 2001.

Eriksson K.F. and Lindgärde F. Prevention of type 2 (non-insulin-dependant) diabetes mellitus by diet and physical exercise. *Diabetologia*, **34**, 891-898, 1991.

Esposito D.L., Mammarella S., Ranieri A., Loggia F.D., Capani F., Consoli A., Mariani-Costantini R., Caramia F.G., Cama A. and Battista P. Deletion of gly723 in the insulin receptor substrate-1 of a patient with noninsulin-dependent diabetes mellitus. *Hum. Mutat.*, **7**, 364-366, 1996.

Etgen G.J., Valasek K.M., Broderick C.L. and Miller A.R. *In vivo* adenoviral delivery of recombinant human protein kinase C-ζ stimulates glucose transport activity in rat skeletal muscle. *J. Biol. Chem.*, **274**, 22139-22142, 1999.

Evans J.C., Frayling T.M., Cassell P.G., Saker P.J., Hitman G.A., Walker M., Levy J.C., O'Rahilly S., Rao P.V.S., Bennett A.J., Jones E.C., Menzel S., Prestwich P., Simecek N., Wishart M., Dhillon R., Fletcher C., Millward A., Demaine A., Wilkin T., Horikawa Y., Cox N.J., Bell G.I., Ellard S.,

McCarthy M.I. and Hattersley A.T. Studies of association between the gene for calpain 10 and type 2 diabetes mellitus in the United Kingdom. *Am. J. Hum. Genet.*, **69**, 544-552, 2001.

Expert panel on detection, evaluation and treatment of high blood cholesterol in adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA.*, **19**, 2486-2497, 2001.

Faber O.K. and Damsgaard E.M. Insulin secretion in type II diabetes. *Acta Endocrinologica*, **262**, 47-50, 1984.

Fang X., Yu S.X., Lu Y., Bast R.C., Woodgett J.R. and Mills G.B. Phosphorylation and inactivation of glycogen synthase kinase 3 by protein kinase A. *Proc. Natl. Acad. Sci. USA.*, **97**, 11960-11965, 2000.

Fang X., Yu S.X., Tanyi J.L., Lu Y., Woodgett J.R. and Mills G.B. Convergence of multiple signaling cascades at glycogen synthase kinase 3: Edg receptor-mediated phosphorylation and inactivation by lysophosphatidic acid through a protein kinase C-dependent intracellular pathway. *Mol. Cell. Biol.*, **22**, 2099-2110, 2002.

Fasshauer M., Klein J., Neumann S., Eszlinger M. and Paschke R. Adiponectin gene expression is inhibited by β-adrenergic stimulation via protein kinase A in 3T3-L1 adipocytes. *FEBS. Lett.*, **507**, 142-146, 2001.

Feise R.J. Do multiple outcomes measures require p-value adjustment? *BMC Medical Research Methodology*, **2**, 8-11, 2002.

Fernández-Real J.M., Gutierrez C., Ricart W., Casamitjana R., Fernández-Castañer M., Vendrell J., Richart C. and Soler J. The TNF-α gene *Nco* I polymorphism influences the relationship among insulin resistance, percent body fat and increased serum leptin levels. *Diabetes*, **46**, 1468-1472, 1997.

Flierl A., Reichmann H. and Seibel P. Pathophysiology of the MELAS 3243 transition mutation. *J. Biol. Chem.*, **272**, 27189-27196, 1997.

Flint J., Harding R.M., Clegg J.B. and Boyce A.J. Why are some genetic diseases common? *Hum Genet.*, **91**, 91-117, 1993.

Formiguera X. and Cantón A. Obesity: epidemiology and clinical aspects. *Best Pract. Res. Clin. Gastroenterol.*, **18**, 1125-1146, 2004.

Fossey S.C., Price J.A., Brewer C.C., Freedman B.I. and Bowden D.W. Genomic characterisation of RACK7. *Am. J. Hum. Genet.*, **63**, A1027, 1998.

Franke T.F., Kaplan D.R., Cantley L.C. and Toker A. Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. *Science*, **275**, 665-668, 1997.

Freidenberg G.R., Henry R.R., Klein H.H., Reichart D.R. and Olefsky J.M. Decreased kinase activity of insulin receptors from adipocytes of non-insulin-dependent diabetic patients. *J. Clin. Invest.*, **79**, 240-250, 1987.

Friedman J.M. Obesity in the new millennium. *Nature*, **404**, 632-634, 2000.

Fritsche A., Madaus A., Renn W., Tschritter O., Teigeler A., Weisser M., Maerker E., Machicao F., Häring H. and Stumvoll M. The prevalent Gly1057Asp polymorphism in the insulin receptor substrate-2 gene is not associated with impaired insulin secretion. *J. Clin. Endocrinol. Metab.* **86**, 4822-4825, 2001.

Froguel P., Vaxillaire M., Sun F., Velho G., Zouali H., Butel M.O., Lesage S., Vionnet N., Clement K., Fougerousse F., Tanizawa Y., Weissenbach J., Beckmann J.S., Lathrop G.M., Passa P., Permutt M.A. and Cohen D. Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature*, **356**, 162-164, 1992.

Froguel P., Zouali H., Vionnet N., Velho G., Vaxillaire M., Sun F., Lesage S., Stoffel M., Takeda J., Passa P., Permutt M.A., Beckmann J.S., Bell G.I. and Cohen D. Familial hyperglycemia due to mutations in glucokinase: definition of a subtype of diabetes mellitus. *N. Engl. J. Med.*, **328**, 697-702, 1993.

Froguel P. and Velho G. Molecular genetics of maturity-onset diabetes of the young. *Trends Endocrinol. Metab.*, **10**, 142-145, 1999.

Fullerton S.M., Bartoszewicz A., Ybazeta G., Horikawa Y., Bell G.I., Kidd K.K., Cox N.J., Hudson R.R. and Di Rienzo A. Geographic and haplotype structure of candidate type 2 diabetes-susceptibility variants at the Calpain-10 locus. *Am. J. Hum. Genet.*, **70**, 1096-1106, 2002.

Furuta H., Furuta M., Sanke T., Ekawa K., Hanabusa T., Nishi M., Sasaki H. and Nanjo K. Nonsense and missense mutations in the human hepatocyte nuclear factor-1-β gene (TCF2) and their relation to type 2 diabetes in Japanese. *J. Clin. Endocrinol. Metab.*, **87**, 3859-3863, 2002.

Garant M.J., Kao W.H.L., Brancati F., Coresh J., Rami T.M., Hanis C.L., Boerwinkle E. and Shuldiner A.R. SNP43 of CAPN10 and the risk of type 2 diabetes in African-Americans. *Diabetes*, **51**, 231-237, 2002.

Garvey W.T., Maianu L., Hancock J.A., Golichowski A.M. and Baron A. Gene expression of GLUT4 in skeletal muscle from insulin resistant patients with obesity, IGT, GDM, and NIDDM. *Diabetes*, **41**, 465-475, 1992.

Georgopoulos A., Aras O. and Tsai M.Y. Codon-54 polymorphism of the fatty acid-binding protein 2 gene is associated with elevation of fasting and postprandial triglyceride in type 2 diabetes. *J. Clin. Endocrinol. Metab.*, **85**, 3155-3160, 2000.

Ghilardi N., Ziegler S., Wiestner A., Stoffel R., Heim M.H. and Skoda R.C. Defective STAT signaling by the leptin receptor in diabetic mice. *Proc. Natl. Acad. Sci. USA.*, **93**, 6231-6235, 1996.

Ghosh S., Watanabe R.M., Hauser E.R., Valle T., Magnuson V.L., Erdos M.R., Langefeld C.D., Balow J., Ally D.S., Kohtamaki K., Chines P., Birznieks G., Kaleta H-S., Musick A., Te C., Tannenbaum J., Eldridge W., Shapiro S., Martin C., Witt A., So A., Chang J., Shurtleff B., Porter R., Kudelko K., Unni A., Segal L., Sharaf R., Blaschak-Harvan J., Eriksson J., Tenkula T., Vidgren G., Ehnholm C., Tuomilehto-Wolf E., Hagopian W., Buchanan T.A., Tuomilehto J., Bergman R.N., Collins F.C. and Boehnke M. Type 2 diabetes: Evidence for linkage on chromosome 20 in 716 Finnish affected sib pairs. *Proc. Natl. Acad. Sci. USA.*, **96**, 2198-2203, 1999.

Giles R.E., Blanc H., Cann H.M. and Wallace D.C. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA.*, **77**, 6715-6719, 1980.

Gong D.-W., Bi S., Pratley R.E. and Weintraub B.D. Genomic structure and promoter analysis of the human *obese* gene. *J. Biol. Chem.*, **271**, 3971-3974, 1996.

Gonzalez E., Dhanda R., Bamshad M., Mummidi S., Geevarghese R., Catano G., Anderson S.A., Walter E.A., Stephan K.T., Hammer M.F., Mangano A., Sen L., Clark R.A., Ahuja S.S., Dolan M.J. and Ahuja S.K. Global survey of genetic variation in CCR5, RANTES and MIP-1α: Impact on the epidemiology of the HIV-1 pandemic. *Proc. Natl. Acad. Sci. USA.*, **98**, 5199-5204, 2001.

Gordon-Larsen P., Harris K.M., Ward D.S. and Popkin B.M. Acculturation and overweight-related behaviors among Hispanic immigrants to the US: the National Longitudinal Study of Adolescent Health. *Soc. Sci. Med.*, **57**, 2023-2034, 2003.

Görgens H., Schwarz P., Schulze J. and Schackert H.K. LightCycler assay in the analysis of haplotypes of the type 2 diabetes susceptibility gene CAPN10. *Clin. Chem.*, **49**, 1405-1408, 2003.

Gould G.W., Jess T.J., Andrews G.C., Herbst J.J., Plevin R.J. and Gibbs E.M. Evidence for a role of phosphatidylinositol 3-kinase in the regulation of glucose transport in *Xenopus* oocytes. *J. Biol. Chem.*, **269**, 26622-26625, 1994.

Gray I.C., Campbell D.A. and Spurr N.K. Single nucleotide polymorphisms as tools in human genetics. *Hum. Mol. Genet.*, **9**, 2403-2408, 2000.

Groop L.C., Kankuri M., Schalin-Jantti C., Ekstrand A., Nikula-Ijas P., Widen E., Kuismanen E., Eriksson J., Franssila-Kallunki A., Saloranta C. and Koskimies S. Association between polymorphism of the glycogen synthase gene and non-insulin-dependent diabetes mellitus. *N. Engl. J. Med.*, **328**, 10-14, 1993.

Hales C.N., Barker D.J.P., Clark P.M.S., Cox L.J., Fall C., Osmond C. and Winter P.D. Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ.*, **303**, 1019-1022, 1991.

Hales C.N. and Barker D.J.P. Type 2 (non-insulin-dependant) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia*, **35**, 595-601, 1992.

Hani E.H., Suaud L., Boutin P., Chevre J.-C., Durand E., Philippi A., Demenais F., Vionnet N., Furuta H., Velho G., Bell G. I., Laine B. and Froguel P. A missense mutation in hepatocyte nuclear factor-4-α, resulting in a reduced transactivation activity, in human late-onset non-insulin-dependent diabetes mellitus. *J. Clin. Invest.*, **101**, 521-526, 1998.

Hani E.H., Stoffers D.A., Chèvre J-C., Durand E., Stanojevic V., Dina C., Habener J.F. and Froguel P. Defective mutations in the insulin promoter factor-1 (IPF-1) gene in late-onset type 2 diabetes mellitus. *J. Clin. Invest.*, **104**, R41-R48, 1999.

Hanis C.L., Boerwinkle E., Chakraborty R., Ellsworth D.L., Concannon P., Stirling B., Morrison V.A., Wapelhorst B., Spielman R.S., Gogolin-Ewens K.J., Shepard J.M., Williams S.R., Risch N., Hinds D., Iwasaki N., Ogata M., Omori Y., Petzold C., Rietzsch H., Schröder H.-E., Schulze J., Cox N.J., Menzel S., Boriraj V.V., Chen X., Lim L.R., Lindner T., Mereu L.E., Wang Y.-Q., Xiang K., Yamagata K., Yang Y. and Bell G.I. A genome-wide search for human non-insulin-dependant (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat. Genet.*, **13**, 161-166, 1996.

Hansen L., Arden K.C., Rasmussen S.B., Viars C.S., Vestergaard H., Hansen T., Moller A.M., Woodgett J.R. and Pedersen O. Chromosomal mapping and mutational analysis of the coding region of the glycogen synthase kinase-3α and β isoforms in patients with NIDDM. *Diabetologia*, **40**, 940-946, 1997.

Hansen B.C., Jen K-L.C., Pek S.B. and Wolfe R.A. Rapid oscillations in plasma insulin, glucagon and glucose in obese and normal weight humans. *J. Clin. Endocrinol. Metab.*, **54**, 785-792,1982.

Hara K., Boutin P., Mori Y., Tobe K., Dina C., Yasuda K., Yamauchi T., Otabe S., Okada T., Eto K., Kadowaki H., Hagura R., Akanuma Y., Yazaki Y., Nagai R., Taniyama M., Matsubara K., Yoda M., Nakano Y., Kimura S., Tomita M., Kimura S., Ito C., Froguel P. and Kadowaki T. Genetic variation in the gene encoding adiponectin is associated with an increased risk of type 2 diabetes in the Japanese population. *Diabetes,* **51**, 536-540, 2002.

Hardy G.H. Mendelian proportions in a mixed population. *Science*, **28**, 49-50, 1908.

Haruta T., Uno T., Kawahara J., Takano A., Egawa K., Sharma P.M., Olefsky J.M. and Kobayashi M. A rapamycin-sensitive pathway down-regulates insulin signaling via phosphorylation and proteasomal degradation of insulin receptor substrate-1. *Mol. Endocrinol.*, **14**, 783-794, 2000.

Hayashi K. and Yandell D.W. How sensitive is PCR-SSCP? *Hum. Mutat.*, **2**, 338-346, 1993.

He Y., Chen H., Quon M.J. and Reitman M. The mouse 'obese' gene: genomic organization, promoter activity, and activation by CCAAT/enhancer-binding protein-α. *J. Biol. Chem.*, **270**, 28887-28891, 1995.

Hegele R.A., Harris S.B., Hanley A.J.G., Sadikian S., Connelly P.W. and Zinman B. Genetic variation of intestinal fatty acid-binding protein associated with variation in body mass in aboriginal Canadians. *J. Clin. Endocrinol. Metab.*, **81**, 4334-4337, 1996.

Hegele R.A., Harris S.B., Zinman B., Wang J., Cao H., Hanley A.J.G., Tsui L.-C. and Scherer S.W. Variation in the AU(AT)-rich element within the 3'-untranslated region of PPP1R3 is associated with variation in plasma glucose in aboriginal Canadians. *J. Clin. Endocrinol. Metab.*, **83**, 3980-3983, 1998.

Hegele R.A., Cao H., Harris S.B., Hanley A.J.G. and Zinman B. The hepatic nuclear factor-1α G319S variant is associated with early onset type 2 diabetes in Canadian Oji-Cree. *J. Clin. Endocrinol. Metab.*, **84**, 1077-1082, 1999.

Helps N.R., Street A.J., Elledge S.J. and Cohen P.T.W. Cloning of the complete coding region for human protein phosphatase inhibitor 2 using the two hybrid system and expression of inhibitor 2 in *E. coli*. *FEBS. Lett.*, **340**, 93-98, 1994.

Hess J.F., Parisi M.A., Bennett J.L. and Clayton D.A. Impairment of mitochondrial transcription termination by a point mutation associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature*, **351**, 236-239, 1991.

Himsworth H.P. Diabetes mellitus: its differentiation into insulin-sensitive and insulin insensitive types. *Lancet*, **1**, 127-130, 1936.

Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M., Hinokio Y., Lindner T.H., Mashima H., Schwarz P.E.H., del Bosque-Plata L., Horikawa Y., Oda Y., Yoshiuchi I., Colilla S., Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schulze J., Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L. and Bell G.I. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.*, **26**, 163-175, 2000.

Hotamisligil G.S. and Spiegelman B.M. Tumor necrosis factor α: a key component of the obesity-diabetes link. *Diabetes*, **43**, 1271-1278, 1994.

Hotamisligil G.S., Arner P., Caro J.F., Atkinson R.L. and Spiegelman B.M. Increased adipose tissue expression of tumor necrosis factor-α in human obesity and insulin resistance. *J. Clin. Invest.*, **95**, 2409-2415, 1995.

Hother-Nielsen O., Schmitz O., Andersen P.H., Beck-Nielsen H. and Pederson O. Metformin improves peripheral but not hepatic insulin action in obese patients with type II diabetes. *Acta Endocrinologica*, **120**, 257-265, 1989.

Hribal M.L., Federici M., Porzio O., Lauro D., Borboni P., Accili D., Lauro R. and Sesti G. The gly-to-arg^{972} amino acid polymorphism in insulin receptor substrate-1 affects glucose metabolism in skeletal muscle cells. *J. Clin. Endocrinol. Metab.*, **85**, 2004-2013, 2000.

Hu P., Mondino A., Skolnik E.Y. and Schlessinger J. Cloning of a novel, ubiquitously expressed human phosphatidylinositol 3'-kinase and identification of its binding site on p85. *Mol. Cell. Biol.*, **13**, 7677-7688, 1993.

Hua Q.-X., Zhao M., Narayana N., Nakagawa S.H., Jia W. and Weiss M.A. Diabetes-associated mutations in a β-cell transcription factor destabilize an antiparallel "mini-zipper" in a dimerisation interface. *Proc. Natl. Acad. Sci. USA.*, **97**, 1999-2004, 2000.

Huang F.L. and Glinsmann W.H. Separation and characterization of two phosphorylase phosphatase inhibitors from rabbit skeletal muscle. *Eur. J. Biochem.*, **70**, 419-426, 1976.

Huang X., Vaag A., Hansson M., Weng J., Laurila E. and Groop L. Impaired insulin-stimulated expression of the glycogen synthase gene in skeletal muscle of type 2 diabetic patients is acquired rather than inherited. *J. Clin. Endocrinol. Metab.*, **85**, 1584-1590, 2000.

Hubbard S.R. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J.*, **16**, 5572-5581,1997.

Hughes K., Nikolakaki E., Plyte S.E., Totty N.F. and Woodgett J.R. Modulation of the glycogen synthase kinase-3 family by tyrosine phosphorylation. *EMBO J.*, **12**, 803-808, 1993.

Hughes R.I. and Aitman T.J. Genetics of the metabolic syndrome and implications for therapy. *International Congress Series*, **1262**, 224-229, 2004.

Humphries S.E., Luong L.A., Ogg M.S., Hawe E. and Miller G.J. The interleukin-6 -174 G/C promoter polymorphism is associated with risk of coronary heart disease and systolic blood pressure in healthy men. *Eur. Heart. J.*, **22**, 2243-2252, 2001.

Isse N., Ogawa Y., Tamura N., Masuzaki H., Mori K., Okazaki T., Satoh N., Shigemoto M., Yoshimasa Y., Nishi S., Hosoda K., Inazawa J. and Nakao K. Structural organization and chromosomal assignment of the human *obese* gene. *J. Biol. Chem.*, **270**, 27728-27733, 1995.

Johnson M.J., Wallace D.C. Ferris S.D., Rattazzi M.C. and Cavalli-Sforza L.L. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.*, **19**, 255-271, 1983.

Jonsson J., Carlsson L., Edlund T. and Edlund H. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature*, **371**, 606-609, 1994.

Kadowaki T., Kadowaki H., Rechler M.M., Serrano-Rios M., Roth J., Gorden P. and Taylor S.I. Five mutant alleles of the insulin receptor gene in patients with genetic forms of insulin resistance. *J. Clin. Invest.*, **86**, 254-264, 1990a.

Kadowaki T., Kadowaki H. and Taylor S.I. A nonsense mutation causing decreased levels of insulin receptor mRNA: detection by a simplified technique for direct sequencing of genomic DNA amplified by the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA.*, **87**, 658-662, 1990b.

Kadowaki T., Kadowaki H., Accili D. and Taylor S.I. Substitution of lysine for asparagine at position 15 in the α-subunit of the human insulin receptor: a mutation that impairs transport of receptors to the cell surface and decreases the affinity of insulin binding. *J. Biol. Chem.*, **265**, 19143-19150, 1990c.

Kadowaki T., Kadowaki H., Accili D., Yazaki Y. and Taylor S.I. Substitution of arginine for histidine at position 209 in the α-subunit of the human insulin receptor: a mutation that impairs receptor dimerization and transport of receptors to the cell surface. *J. Biol. Chem.*, **266**, 19143-19150, 1991.

Kahn S.E., Prigeon R.L., McCulloch D.K., Boyko E.J., Bergman R.N., Schwartz M.W., Neifing J.L., Ward W.K., Beard J.C., Palmer J.P. and Porte D. Quantification of the relationship between insulin sensitivity and β-cell function in human subjects. *Diabetes*, **42**, 1663-1672, 1993.

Kellogg D.E., Rybalkin I., Chen S., Mukhamedova N., Vlasik T., Siebert P.D. and Chenchik A. TaqStart Antibody™: 'Hot-Start' PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *BioTechniques.*, **16**, 1134-1137, 1994.

Kido Y., Burks D.J., Withers D., Bruning J.C., Kahn C.R., White M.F. and Accili D. Tissue-specific insulin resistance in mice with mutations in the insulin receptor, IRS-1 and IRS-2. *J. Clin. Invest.*, **105**, 199-205, 2000.

King H., Aubert R.E. and Herman W.H. Global burden of diabetes, 1995-2025: Prevalence, numerical estimates and projections. *Diabetes Care*, **21**, 1414-1431, 1998.

Kingston M.E. and Skoog W.C. Maintenance of basal insulin secretion in severe non-insulin dependant diabetes. *Diabetes Care*, **9**, 232-235, 1986.

Kissebah A.H., Sonnenberg G.E., Myklebust J., Goldstein M., Broman K., James R.G., Marks J.A., Krakower G.R., Jacob H.I., Weber J., Martin L., Blangero J. and Comuzzie A.G. Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc. Natl. Acad. Sci. USA.*, **97**, 14478-14483, 2000.

Knowler W.C., Narayan K.M.V., Hanson R.L., Nelson R.G., Bennett P.H., Tuomilehto J., Schersten B. and Pettitt D.J. Preventing non-insulin-dependant diabetes. *Diabetes*, **44**, 483-488, 1995.

Kobayashi T., Nakanishi K., Nakase H., Kajio H., Okubo M., Murase T. and Kosaka K. *In situ* characterization of islets in diabetes with a mitochondrial DNA mutation at nucleotide position 3243. *Diabetes*, **46**, 1567-1571, 1997.

Kohn A.D., Summers S.A., Birnbaum M.J. and Roth R.A. Expression of a constitutively active Akt ser/thr kinase in 3T3-L1 adipocytes stimulates glucose uptake and glucose transporter 4 translocation. *J. Biol. Chem.*, **271**, 31372-31378, 1996a.

Kohn A.D., Takeuchi F. and Roth R.A. Akt, a pleckstrin homology domain containing kinase, is activated primarily by phosphorylation. *J. Biol. Chem.*, **271**, 21920-21926, 1996b.

Koistinen H.A., Koivisto V.A., Andersson S., Karonen S.-L., Kontula K., Oksanen L. and Teramo K.A. Leptin concentration in cord blood correlates with intrauterine growth. *J. Clin. Endocrinol. Metab.*, **82**, 3328-3330, 1997.

Kolterman O.G., Insel J., Saekow M. and Olefsky J.M. Mechanisms of insulin resistance in human obesity. *J. Clin. Invest.*, **65**, 1272-1284, 1980.

Kondo H., Shimomura I., Matsukawa Y., Kumada M., Takahashi M., Matsuda M., Ouchi N., Kihara S., Kawamoto T., Sumitsuji S., Funahashi T. and Matsuzawa Y. Association of adiponectin mutation with type 2 diabetes. *Diabetes*, **51**, 2325-2328, 2002.

Kulkarni R.N., Winnay J.N., Daniels M., Bruning J.C., Flier S.N., Hanahan D. and Kahn C.R. Altered function of insulin receptor substrate-1-deficient mouse islets and cultured β-cell lines. *J. Clin. Invest.*, **104**, R69-R75, 1999.

Kuo C.J., Conley P.B., Hsieh C.-L., Francke U. and Crabtree G.R. Molecular cloning, functional expression, and chromosomal localization of mouse hepatocyte nuclear factor 1. *Proc. Natl. Acad. Sci. USA.*, **87**, 9838-9842, 1990.

Laakso M., Malkki M., Kekalainen P., Kuusisto J. and Deeb S.S. Insulin receptor substrate-1 variants in non-insulin-dependent diabetes. *J. Clin. Invest.*, **94**, 1141-1146, 1994.

Lausen J., Thomas H., Lemm I., Bulman M., Borgschulze M., Lingott A., Hattersley A.T. and Ryffel G.U. Naturally occurring mutations in the human HNF4-α gene impair the function of the transcription factor to a varying degree. *Nucleic Acids Res.*, **28**, 430-437, 2000.

Le Good J.A., Ziegler W.H., Parekh D.B., Alessi D.R., Cohen P. and Parker P.J. Protein kinase C isotypes controlled by phosphoinositide 3-kinase through the protein kinase PDK-1. *Science*, **281**, 2042-2045, 1998.

Lev-Ran A., Sprecher E., Yerushalmy Y., Schindel B. and Kisch E.S. Homogeneity of the age at diagnosis in sibs with type 2 diabetes: Implications for sib-pair analysis. *Am. J. Med. Genet.*, **91**, 91-94, 2000.

Lindner T. H., Njølstad P.R., Horikawa Y., Bostad L., Bell G. I. and Søvik O. A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1-β. *Hum. Mol. Genet.*, **8**, 2001-2008, 1999.

Long J.C., Williams R.C. and Urbanek M. An E-M algorithm and testing strategy for multiple-locus haplotypes. *Am. J. Hum. Genet.*, **56**, 799-810, 1995.

Macfarlane W.M., Frayling T.M., Ellard S., Evans J.C., Allen L.I.S., Bulman M.P., Ayres S., Shepherd M., Clark P., Millward A., Demaine A., Wilkin T., Docherty K. and Hattersley A.T. Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. *J. Clin. Invest.*, **104**, R33-R39, 1999.

Maffei M., Fei H., Lee G.-H., Dani C., Leroy P., Zhang Y., Proenca R., Negrel R., Ailhaud G. and Friedman J.M. Increased expression in adipocytes of ob RNA in mice with lesions of the hypothalamus and with mutations at the db locus. *Proc. Natl. Acad. Sci. USA.*, **92**, 6957-6960, 1995.

Magnuson M.A. Glucokinase gene structure: functional implications of molecular genetic studies. *Diabetes*, **39**, 523-527, 1990.

Mahtani M.M., Widen E., Lehto M., Thomas J., McCarthy M., Brayer J., Bryant B., Chan G., Daly M., Forsblom C., Kanninen T., Kirby A., Kruglyak L., Munnelly K., Parkkonen M., Reeve-Daly M.P., Weaver A., Brettin T., Duyk G., Lander E.S. and Groop L.C. Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat. Genet.*, **14**, 90-94, 1996.

Majer M., Mott D.M., Mochizuki H., Rowles J.C., Pedersen O., Knowler W.C., Bogardus C. and Prochazka M. Association of the glycogen synthase locus on 19q13 with NIDDM in Pima Indians. *Diabetologia*, **39**, 314-321, 1996.

Malecki M.T., Moczulski D.K., Klupa T., Wanic K., Cyganek K., Frey J. and Sieradzki J. Homozygous combination of calpain 10 gene haplotypes is associated with type 2 diabetes mellitus in a Polish population. *Eur. J. Endocrinol.*, **146**, 695-699, 2002.

Mammarella S., Romano F., Di Valerio A., Creati B., Esposito D.L., Palmirotta R., Capani F., Vitullo P., Volpe G., Battista P., Loggia F.D., Mariani-Constantini R. and Carna A. Interaction between the G1057D variant of IRS-2 and overweight in the pathogenesis of type 2 diabetes. *Hum. Mol. Genet.* **9**, 2517-2521, 2000.

Matschinsky F.M. Glucokinase as glucose sensor and metabolic signal generator in pancreatic β-cells and hepatocytes. *Diabetes*, **39**, 647-652, 1990.

Matsubara M., Maruoka S. and Katayose S. Inverse relationship between plasma adiponectin and leptin concentrations in normal-weight and obese woman. *Eur. J. Endocrin.*, **147**, 173-180, 2002.

Matsuda J., Yokota I., Iida M., Murakami T., Naito E., Ito M., Shima K. and Kuroda Y. Serum leptin concentration in cord blood: relationship to birth weight and gender. *J. Clin. Endocrinol. Metab.*, **82**, 1642-1644, 1997.

Mauvais-Jarvis F., Ueki K., Fruman D.A., Hirshman M.F., Sakamoto K., Goodyear L.J., Iannacone M., Accili D., Cantley L.C. and Kahn C.R. Reduced expression of the murine p85α subunit of the phosphatidylinositide 3-kinase improves insulin signalling and ameliorates diabetes. *J. Clin. Invest.*, **109**, 141-149, 2002.

Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S., Brandon M., Easley K., Chen E., Brown M.D., Sukernik R.I., Olckers A. and Wallace D.C. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA.*, **100**, 171-176, 2003.

McGeer P.L., Schulzer M. and McGeer E.G. Arthritis and antiinflammatory agents as possible protective factors for Alzheimer disease: a review of seventeen epidemiological studies. *Neurology*, **47**, 425-432, 1996.

Modan M., Halkin H., Almog S., Lusky A., Eshkol A., Shefi M., Shitrit A. and Fuchs A. Hyperinsulinemia: a link between hypertension, obesity, and glucose intolerance. *J. Clin. Invest.*, **75**, 809-817, 1985.

Moller D.E. and Flier J.S. Detection of an alteration in the insulin-receptor gene in a patient with insulin resistance, acanthosis nigricans, and polycystic ovary syndrome (Type A insulin resistance). *N. Engl. J. Med.*, **319**, 1526-1529, 1988.

Moller D.E., Benecke H. and Flier J.S. Biologic activities of naturally occurring human insulin receptor mutations: evidence that metabolic effects of insulin can be mediated by a kinase-deficient insulin receptor mutant. *J. Biol. Chem.*, **266**, 10995-11001, 1991.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G. and Erlich H. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *Cold Spring Harbour Symp. Quant. Biol.*, **L1**, 263-273, 1986.

Mullis K.B. and Faloona F.A. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.*, **155**, 335-350, 1987.

Myers M.G., Backer J.M., Sun X.J., Shoelson S., Hu P., Schlessinger J., Yoakim M., Schaffhausen B. and White M.F. IRS-1 activates phosphatidylinositol 3'-kinase by associating with *src* homology 2 domains of p85. *Proc. Natl. Acad. Sci. USA.*, **89**, 10350-10354, 1992.

Nakanishi S., Okubo M., Yoneda M., Jitsuiki K., Yamane K. and Kohno N. A comparison between Japanese-Americans living in Hawaii and Los Angeles and native Japanese: the impact of lifestyle westernization on diabetes mellitus. *Biomed. Pharmacother.*, **58**, 571-577, 2004.

Napoli C., Lerman L.O., de Nigris F., Loscalzo J. and Ignarro L.J. Glycoxidized low-density lipoprotein downregulates endothelial nitric oxide synthase in human coronary cells. *J. Am. Coll. Cardiol.*, **40**, 1515-1522, 2002.

Narang S.A., Brousseau R., Georges F., Michniewicz J., Prefontaine G., Stawinski J. and Sung W. The human preproinsulin gene: synthesis, cloning, gene modification and expression studies. *Can. J. Biochem.*, **62**, 209-216, 1984.

Nauck M., März W., Hoffmann M.M., Nagel D., Boehm B.O. and Winkelmann B.R. The G(-30)A polymorphism in the promoter of the glucokinase gene is associated with angiographic coronary artery disease, type 2 diabetes mellitus and glucose metabolism (abstract). The XIIIth International Symposium on Atherosclerosis, Kyoto, Japan, 2003.

Naya F.J., Huang H.-P., Qiu Y., Mutoh H., DeMayo F.J., Leiter A.B. and Tsai M.-J. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes Dev.*, **11**, 2323-2334, 1997.

Nishigori H., Yamada S., Kohama T., Tomura H., Sho K., Horikawa Y., Bell G. I., Takeuchi T. and Takeda J. Frameshift mutation, A263fsinsGG, in the hepatocyte nuclear factor-1-β gene associated with diabetes and renal dysfunction. *Diabetes*, **47**, 1354-1355, 1998.

Nishiyama M., Inazawa J., Ariyama T., Nakamura Y., Matsufuji S., Furusaka A., Tanaka T., Hayashi S. and Wands J.R. The human insulin receptor substrate-1 gene (IRS-1) is localized on 2q36. *Genomics*, **20**, 139-141, 1994.

Noguchi T., Matozaki T., Horita K., Fujioka Y. and Kasuga M. Role of SH-PTP2, a protein-tyrosine phosphatase with *src* homology 2 domains, in insulin-stimulated ras activation. *Mol. Cell. Biol.*, **14**, 6674-6682, 1994.

Nolan J.J., Ludvik B., Beerdsen P., Joyce M. and Olefsky J. Improvement in glucose tolerance and insulin resistance in obese subjects treated with troglitazone. *N. Engl. J. Med.*, **331**, 1188-1193, 1994.

Ogihara T., Isobe T., Ichimura T., Taoka M., Funaki M., Sakoda H., Onishi Y., Inukai K., Anai M., Fukushima Y., Kikuchi M., Yazaki Y., Oka Y. and Asano T. 14-3-3 protein binds to insulin receptor substrate-1, one of the binding sites of which is in the phosphotyrosine binding domain. *J. Biol. Chem.*, **272**, 25267-25274, 1997.

Ohlsson H., Karlsson K. and Edlund T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J.*, **12**, 4251-4259, 1993.

Olckers A., Prosser D., Wallace D.C., Brown M.D. and Smuts I. Mitochondrial myopathies: a South African perspective (abstract). *Am. J. Hum. Genet.*, **69**, 490, 2001.

Olefsky J.M., Farquhar J.W. and Reavan G.M. Reappraisal of the role of insulin in hypertriglyceridemia. *Am. J. Med.*, **57**, 551-560, 1974.

Omar M.A.K., Seedat M.A., Motala A.A., Dyer R.B. and Becker P. The prevalence of diabetes mellitus and impaired glucose tolerance in a group of urban South African blacks. *S. Afr. Med. J.*, **83**, 641-643, 1993.

Orho M., Carlsson M., Kanninen T. and Groop L.C. Polymorphism at the *rad* gene is not associated with NIDDM in Finns. *Diabetes*, **45**, 429-433, 1996.

Ort T., Voronov S., Guo J., Zawalich K., Froehner S.C., Zawalich W. and Solimena M. Dephosphorylation of β2-syntrophin and Ca^{2+}/μ-calpain-mediated cleavage of ICA512 upon stimulation of insulin secretion. *EMBO J.*, **20**, 4013-4023, 2001.

Ouchi N., Kihara S., Arita Y., Maeda K., Kuriyama H., Okamoto Y., Hotta K., Nishida M., Takahashi M., Nakamura T., Yamashita S., Funahashi T. and Matsuzawa Y. Novel modulator for endothelial adhesion molecules adipocyte-derived plasma protein adiponectin. *Circulation*, **100**, 2473-2476, 1999.

Ouchi N., Kihara S., Arita Y., Okamoto Y., Maeda K., Kuriyama H., Hotta K., Nishida M., Takahashi M., Muraguchi M., Ohmoto Y., Nakamura T., Yamashita S., Funahashi T. and Matsuzawa Y. Adiponectin, an adipocyte derived plasma protein, inhibits endothelial NF-κB signalling through a cAMP-dependant pathway. *Circulation*, **102**, 1296-1301, 2000.

Patti M.-E., Sun X.-J., Bruening J.C., Araki E., Lipes M.A., White M.F. and Kahn C.R. 4PS/Insulin receptor substrate (IRS)-2 is the alternative substrate of the insulin receptor in IRS-1 deficient mice. *J. Biol. Chem.*, **270**, 24670-24673, 1995.

Pelleymounter M.A., Cullen M.J., Baker M.B., Hecht R., Winters D., Boone T. and Collins F. Effects of the obese gene product on body weight regulation in ob/ob mice. *Science*, **269**, 540-543, 1995.

Pennisi E. Searching for the genome's second code. *Science*, **306**, 632-635, 2004.

Permana P.A. and Mott D.M. Genetic analysis of human type 1 protein phosphatase inhibitor 2 in insulin resistant Pima Indians. *Genomics*, **41**, 110-114, 1997.

Pickup J.C., Chusney G.D., Thomas S.M. and Burt D. Plasma interleukin-6, tumour necrosis factor α and blood cytokine production in type 2 diabetes. *Life Sci.*, **67**, 291-300, 2000.

Polonsky K.S., Sturis J. and Bell G.I. Non-insulin-dependant diabetes mellitus-a genetically programmed failure of the beta-cell to compensate for insulin resistance. *N. Engl. J. Med.*, **334**, 777-783, 1996.

Porzio O., Federici M., Hribal M.L., Lauro D., Accili D., Lauro R., Borboni P. and Sesti G. The gly^{972}-to-arg amino acid polymorphism in IRS-1 impairs insulin secretion in pancreatic β cells. *J. Clin. Invest.*, **104**, 357-364, 1999.

Price J.A., Brewer C.S., Howard T.D., Fossey S.C., Sale M.M., Ji L., Krolewski A.S. and Bowden D.W. A physical map of the 20q12-q13.1 region associated with type 2 diabetes. *Genomics*, **62**, 208-215, 1999.

Prochazka M., Lillioja S., Tait J.F., Knowler W.C., Mott D.M., Spraul M., Bennett P.H. and Bogardus C. Linkage of chromosomal markers on 4q with a putative gene determining maximal insulin action in Pima Indians. *Diabetes*, **42**, 514-519, 1993.

Randle P.J., Garland P.B., Hales C.N. and Newsholme E.A. The glucose fatty-acid cycle: Its role in insulin sensitivity and the metabolic disturbances of diabetes mellitus. *Lancet*, 785-789, 1963.

Randle P.J. α-Ketoacid dehydrogenase complexes and respiratory fuel utilisation in diabetes. *Diabetologia*, **28**, 479-484, 1985.

Reavan G.M., Lerner R.L., Stern M.P. and Farquhar J.W. Role of insulin in endogenous hypertriglyceridemia. *J. Clin. Invest.*, **46**, 1756-1767, 1967.

Reavan G.M. Role of insulin resistance in human disease. *Diabetes*, **37**, 1595-1607, 1988.

Rhoads R.E. Regulation of eukaryotic protein synthesis by initiation factors. *J. Biol. Chem.*, **268**, 3017-3020, 1993.

Rosen O.M., Herrara R., Olowe Y., Petruzzelli L.M. and Cobb M.H. Phosphorylation activates the insulin receptor tyrosine protein kinase. *Proc. Natl. Acad. Sci. USA.*, **80**, 3237-3240, 1983.

Rossetti L. and Giaccari A. Relative contribution of glycogen synthesis and glycolysis to insulin-mediated glucose uptake. *J. Clin. Invest.*, **85**, 1785-1792, 1990.

Saad M.F., Damani S., Gingerich R.L., Riad-Gabriel M.G., Khan A., Boyadjian R., Jinagouda S.D., El-Tawil K., Rude R.K. and Kamdar V. Sexual dimorphism in plasma leptin concentration. *J. Clin. Endocrinol. Metab.*, **82**, 579-584, 1997.

Saito K., Tobe T., Minoshima S., Asakawa S., Sumiya J., Yoda M., Nakano Y., Shimizu N. and Tomita M. Organization of the gene for gelatin-binding protein (GBP28). *Gene*, **229**, 67-73, 1999.

Sambrook J. and Russell D.W. Molecular Cloning: a laboratory manual. Third Edition. Cold Spring Harbour Laboratory Press, A8.20-A8.21, 2001.

Samuels M.L. Statistics for the life sciences. First Edition. Collier MacMillan publishers, 571, 1989.

Savage D.B., Agostini M., Barroso I., Gurnell M., Luan J., Meirhaeghe A., Harding A-H., Ihrke G., Rajanayagam O., Soos M.A., George S., Berger D., Thomas E.L., Bell J.D., Meeran K., Ross R.J., Vidal-Puig A., Wareham N.J., O'Rahilly S., Chatterjee V.K.K. and Schafer A.J. Digenic inheritance of severe insulin resistance in a human pedigree. *Nat. Genet.*, **31**, 379-383, 2002.

Scarlett J.A., Gray R.S., Griffin J., Olefsky J.M. and Kolterman O.G. Insulin treatment reverses the insulin resistance of type II diabetes mellitus. *Diabetes Care*, **5**, 353-363, 1982.

Schutte A.E., van Rooyen J.M., Huisman H.W., Kruger H.S. and de Ridder J.H. Factor analysis of possible risks for hypertension in a black South African population. *J. Hum. Hypertens.*, **17**, 339-348, 2003.

Schwarz P.E.H. Personal communication. 2005.

Schwarz P.E.H., Towers G.W., Görgens H., Fuecker K., Schwanebeck U., Rietzsch H., Fischer S., Julius U., Schulze J., Olckers A., Vasseur F. and Schackert H.K. Haplotypes in the adiponectin promoter region (*APM1*) are associated with increased diabetes risk in a Caucasian population. Submitted for publication in *J. Clin. Endocrinol. Metab.*, 2004.

Segal K.R., Landt M. and Klein S. Relationship between insulin sensitivity and plasma leptin concentration in lean and obese men. *Diabetes*, **45**, 988-991, 1996.

Sharma S., Jhala U.S., Johnson T., Ferreri K., Leonard J. and Montminy M. Hormonal regulation of an islet-specific enhancer in the pancreatic homeobox gene STF-1. *Mol. Cell. Biol.*, **17**, 2598-2604, 1997.

Shaw J.T.E., Lovelock P.K., Kesting J.B., Cardinal J., Duffy D., Wainwright B. and Cameron D.P. Novel susceptibility gene for late-onset NIDDM is localized to human chromosome 12q. *Diabetes*, **47**, 1793-1796, 1998.

Shimabukuro M., Wang M., Zhou Y., Newgard C.B. and Unger R.H. Protection against lipoapoptosis of β cells through leptin-dependant maintenance of Bcl-2 expression. *Proc. Natl. Acad. Sci. USA.*, **95**, 9558-9561, 1998.

Skolnik E.Y., Batzer A., Li N., Lee C.-H., Lowenstein E., Mohammadi M., Margolis B. and Schlessinger J. The function of Grb-2 in linking the insulin receptor to ras signaling pathways. *Science*, **260**, 1953-1955, 1993.

Song Y., Niu T., Manson J.E., Kwiatkowski D.J. and Liu S. Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *Am. J. Hum. Genet.*, **74**, 208-222, 2004.

Sonksen P. and Sonksen J. Insulin: understanding its action in health and disease. *Br. J. Anaesth.*, **85**, 69-79, 2000.

Spranger J., Kroke A., Mööhlig M., Bergmann M.M., Ristow M., Boeing H. and Pfeiffer A.F.H. Adiponectin and protection against type 2 diabetes mellitus. *Lancet*, **361**, 226-228, 2003.

Stephens J.W., Hurel S.J., Cooper J.A., Acharya J., Miller G.J. and Humphries S.E. A common functional variant in the interleukin-6 gene is associated with increased body mass index in subjects with type 2 diabetes mellitus. *Mol. Genet. Metab.*, **82**, 180-186, 2004.

Stoffel M. and Duncan S.A. The maturity-onset diabetes of the young (MODY1) transcription factor HNF4-α regulates expression of genes required for glucose transport and metabolism. *Proc. Natl. Acad. Sci. USA.*, **94**, 13209-13214, 1997.

Stoffel M., Froguel P., Takeda J., Zouali H., Vionnet N., Nishi S., Weber T., Harrison R.W., Pilkis S.J., Lesage S., Vaxillaire M., Velho G., Sun F., Iris F., Passa P., Cohen D. and Bell G.I. Human glucokinase gene: isolation, characterization and identification of two missense mutations linked to early-onset non-insulin-dependent (type 2) diabetes mellitus. *Proc. Natl. Acad. Sci. USA.*, **89**, 7698-7702, 1992.

Stoffel M., Espinosa R., Keller S.R., Lienhard G.E., Le Beau M.M. and Bell G.I. Human insulin receptor substrate-1 gene (IRS1): chromosomal localization to 2q35-q36.1 and identification of a simple tandem repeat DNA polymorphism. *Diabetologia*, **36**, 335-337, 1993.

Stoffel M., Stein R., Wright C.V.E., Espinosa R., Le Beau M.M. and Bell G.I. Localization of human homeodomain transcription factor insulin promoter factor 1 (IPF1) to chromosome band 13q12.1. *Genomics*, **28**, 125-126, 1995.

Stumvoll M., Tschritter O., Fritsche A., Staiger H., Renn W., Weisser M., Machicao F. and Häring H. Association of the T-G polymorphism in adiponectin (exon 2) with obesity and insulin sensitivity. *Diabetes*, **51**, 37-41, 2002.

Sun X.S., Rothenberg P., Kahn C.R., Backer J.M., Araki E., Wilden P.A., Cahill D.A., Goldstein B.J. and White M.F. Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature*, **352**, 73-77, 1991

Surwit R.S., Petro A.E., Parekh P. and Collins S. Low plasma leptin in response to dietary fat in diabetes- and obesity-prone mice. *Diabetes*, **46**, 1516-1520, 1997.

Suzuki K. and Kono T. Evidence that insulin causes translocation of glucose transport activity to the plasma membrane from an intracellular storage site. *Proc. Natl. Acad. Sci. USA.*, **77**, 2542-2545, 1980.

Sweeting M.J., Sutton A.J. and Lambert P.C. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of rare events. 4th Symposium on Systematic Reviews: Pushing the Boundaries. Oxford, United Kingdom, 2002.

Swislocki A.L.M., Chen Y-D.I., Golay A., Chang M.-O. and Reaven G.M. Insulin suppression of plasma-free fatty acid concentration in normal individuals and patients with Type 2 (non-insulin-dependant) diabetes. *Diabetologia*, **30**, 622-626, 1987.

Takano A., Usui I., Haruta T., Kawahara J., Uno T., Iwata M. and Kobayashi M. Mammalian target of rapamycin pathway regulates insulin signaling via subcellular redistribution of insulin receptor substrate 1 and integrates nutritional signals and metabolic signals of insulin. *Mol. Cell. Biol.*, **21**, 5050-5062, 2001.

Tan P.K., Downey T.J., Spitznagel E.L., Xu P., Fu D., Dimitrov D.S., Lempicki R.A., Raaka B.M. and Cam M.C. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res.*, **31**, 5676-5684, 2003.

Tartaglia L.A., Dembski M., Weng X., Deng N., Culpepper J., Devos R., Richards G.J., Campfield L.A., Clark F.T., Deeds J., Muir C., Sanker S., Moriarty A., Moore K.J., Smutko J.S., Mays G.G., Woolf E.A., Monroe C.A. and Tepper R.I. Identification and expression cloning of a leptin receptor, OB-R. *Cell*, **83**, 1263-1271, 1995.

Tattersall R.B. and Fajans S.S. A difference between the inheritance of classical juvenile onset and maturity onset type diabetes of young people. *Diabetes*, **24**, 44-53, 1975.

Taylor S.I. Lilly lecture: molecular mechanisms of insulin resistance: lessons from patients with mutations in the insulin-receptor gene. *Diabetes*, **41**, 1473-1490, 1992.

't Hart L.M., Stolk R.P., Dekker J.M., Nijpels G., Grobbee D.E., Heine R.J. and Maassen J.A. Prevalence of variants in candidate genes for type 2 diabetes mellitus in the Netherlands: the Rotterdam study and the Hoorn study. *J. Clin. Endocrinol. Metab.*, **84**, 1002-1006, 1999.

Thomas S.H.L., Wisher M.H., Brandenburg D. and Sönsken P.H. Insulin action on adipocytes: Evidence that the anti-lipolytic and lipogenic effects of insulin are mediated by the same receptor. *Biochem. J.*, **184**, 355-360, 1979.

Thomas A.W., Edwards A., Sherratt E.J., Majid A., Gagg J. and Alcolado J.C. Molecular scanning of candidate mitochondrial tRNA genes in type 2 (non-insulin dependant) diabetes mellitus. *J. Med. Genet.*, **33**, 253-255, 1996.

Thomas H., Jaschkowitz K., Bulman M., Frayling T.M., Mitchell S.M.S., Roosen S., Lingott-Frieg A., Tack C.J., Ellard S., Ryffel G.U. and Hattersley A.T. A distant upstream promoter of the HNF-4α gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum. Mol. Genet.*, **10**, 2089-2097, 2001.

Torroni A., Brown M.D., Lott M.T., Newman N.J., Wallace D.C. and the Cuba Neuropathy Field Investigation Team. African, Native American and European mitochondrial DNAs in Cubans from Pinar del Rio province and implications for the recent epidemic neuropathy in Cuba. *Hum. Mutat.*, **5**, 310-317, 1995.

Towers G.W., Prosser D.O., Rheeder P., Wallace D.C. and Olckers A. Molecular screening for the mitochondrial A3243G mutation in a black NIDDM South African population. Annual Society of Endocrinology, Metabolism and Diabetes of South Africa Meeting, Cape Town, South Africa, 2002.

Triggs-Raine B.L., Kirkpatrick R.D., Kelly S.L., Norquay L.D., Cattini P.A., Yamagata K., Hanley A.J.G., Zinman B., Harris S.B., Barrett P.H. and Hegele R.A. HNF1α G319S, a transactivation-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community. *Proc. Natl. Acad. Sci. USA.*, **99**, 4614-4619, 2002.

Tsai H-J., Sun G., Weeks D.E., Kaushal R., Wolujewicz M., McGarvey S.T., Tufa J., Viali S. and Deka R. Type 2 diabetes and three calpain-10 gene polymorphisms in Samoans: no evidence of association. *Am. J. Hum. Genet.*, **69**, 1236-1244, 2001.

Unger R.H., Zhou Y. and Orci L. Regulation of fatty acid homeostasis in cells: novel role of leptin. *Proc. Natl. Acad. Sci. USA.*, **96**, 2327-2332, 1999.

Vaag A., Henriksen J.E. and Beck-Nielsen H. Decreased insulin activation of glycogen synthase in skeletal muscle in young nonobese Caucasian first-degree relatives of patients with non-insulin-dependant diabetes mellitus. *J. Clin. Invest.*, **89**, 782-788, 1992.

Valera A., Pujol A., Pelegrin M. and Bosch F. Transgenic mice overexpressing phosphoenolpyruvate carboxykinase develop non-insulin-dependant diabetes mellitus. *Proc. Natl. Acad. Sci. USA.*, **91**, 9151-9154, 1994.

Valverde A.M., Mur C., Pons S., Alvarez A.M., White M.F., Kahn C.R. and Benito M. Association of insulin receptor substrate 1 (IRS-1) Y895 with Grb-2 mediates the insulin signalling involved in IRS-1 deficient brown adipocyte mitogenesis. *Mol. Cell. Biol.*, **21**, 2269-2280, 2001.

Van den Ouweland J.M.W., Lemkes H.H.P.J., Trembath R.C., Ross R., Velho G., Cohen D., Froguel P. and Maassen J.A. Maternally inherited diabetes and deafness is a distinct subtype of diabetes and associates with a single point mutation in the mitochondrial $tRNA^{Leu(UUR)}$ gene. *Diabetes*, **43**, 746-751, 1994.

Vasseur F., Helbecque N., Dina C., Lobbens S., Delannoy V., Gaget S., Boutin P., Vaxillaire M., Lepretre F., Dupont S., Hara K., Clement K., Bihain B., Kadowaki T. and Froguel P. Single-nucleotide polymorphism haplotypes in both the proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum. Mol. Genet.*, **11**, 2607-2614, 2002.

Vaxillaire M., Rouard M., Yamagata K., Oda N., Kaisaki P.J., Boriraj V.V., Chevre J.-C., Boccio V., Cox R.D., Lathrop G.M., Dussoix P., Phillipe J., Timsit J., Charpentier G., Velho G., Bell G.I. and Froguel P. Identification of nine novel mutations in the hepatocyte nuclear factor 1 alpha gene associated with maturity-onset diabetes of the young (MODY3). *Hum. Mol. Genet.*, **6**, 583-586, 1997.

Vestergaard H., Lund S., Larsen F.S., Bjernum O.J. and Pedersen O. Glycogen synthase and phosphofructokinase protein and mRNA levels in skeletal muscle from insulin-resistant patients with non-insulin-dependant diabetes mellitus. *J. Clin. Invest.*, **91**, 2342-2350, 1993.

Vionnet N., Passa P. and Froguel P. Prevalence of mitochondrial gene mutations in families with diabetes mellitus. *Lancet*, **342**, 1429-1430, 1993.

Voliovitch H., Schindler D.G., Hadari Y.R., Taylor S.I., Accili D. and Zick Y. Tyrosine phosphorylation of insulin receptor substrate-1 *in vivo* depends upon the presence of its pleckstrin homology region. *J. Biol. Chem.*, **270**, 18083-18087, 1995.

Waeber G., Delplanque J., Bonny C., Mooser V., Steinmann M., Widmann C., Maillard A., Miklossy J., Dina C., Hani E.H., Vionnet N., Nicod P., Boutin P. and Froguel P. The gene MAPK8IP1, encoding islet-brain-1, is a candidate for type 2 diabetes. *Nat. Genet.*, **24**, 291-295, 2000.

Watson E., Bauer K., Aman R., Weiss G., von Haeseler A. and Pääbo S. mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.*, **59**, 437-444, 1996.

Weedon M.N., Schwarz P.E.H., Horikawa Y., Iwasaki N., Illig T., Holle R., Rathmann W., Selisko T., Schulze J., Owen K.R. Evans J., del Bosque-Plata L., Hitman G., Walker M., Levy J.C., Sampson M., Bell G.I., McCarthy M.I., Hattersley A.T. and Frayling T.M. Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am. J. Hum. Genet.*, **73**, 1208-1212, 2003.

Welsh G.I. and Proud C.G. Glycogen synthase kinase-3 is rapidly inactivated in response to insulin and phosphorylates eukaryotic initiation factor eIF-2B. *Biochem. J.* **294**, 625-629, 1993.

Weyer C., Funahashi T., Tanaka S., Hotta K., Matsuzawa Y., Pratley R.E. and Tataranni P.A. Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *J. Clin. Endocrinol. Metab.*, **86**, 1930-1935, 2001.

Winick J.D., Stoffel M. and Friedman J.M. Identification of microsatellite markers linked to the human leptin receptor gene on chromosome 1. *Genomics*, **36**, 221-222, 1996.

Withers D.J., Gutierrez J.S., Towery H., Burks D.J., Ren J.-M., Previs S., Zhang Y., Bernal D., Pons S., Shulman G.I., Bonner-Weir S. and White M.F. Disruption of IRS-2 causes type 2 diabetes in mice. *Nature*, **391**, 900-904, 1998.

World Health Organisation Consultation. Definition, diagnosis and classification of diabetes mellitus and its complications. Report of a WHO consultation. 1999.

Wolf G., Trüb T., Ottinger E., Groniga L., Lynch A., White M.F., Miyazaki M., Lee J. and Shoelson S.E. PTB domains of IRS-1 and Shc have distinct but overlapping binding specificities. *J. Biol. Chem.*, **270**, 27407-27410, 1995.

Woolf B. On estimating the relation between blood group and disease. *Ann. Hum. Genet.*, **19**, 251-253, 1955.

Yamagata K., Furuta H., Oda N., Kaisaki P.J., Menzel S., Cox N.J., Fajans S.S., Signorini S., Stoffel M. and Bell G.I. Mutations in the hepatocyte nuclear factor-4α gene in maturity-onset diabetes of the young (MODY1). *Nature*, 384, 458-460, 1996a.

Yamagata K., Oda N., Kalsaki P.J., Menzel S., Furuta H., Vaxillaire M., Southam L., Cox R.D., Lathrop G.M., Boriraj V.V., Chen X., Cox N.J., Oda Y., Yano H., Le Beau M.M., Yamada S., Nishigori H., Takeda J., Fajans S.S., Hattersley A.T., Iwasaki N., Hansen T., Pedersen O., Polonsky K.S., Turner R.C., Velho G., Chevre J.-C., Froguel P. and Bell G.I. Mutations in the hepatocyte nuclear factor-1α gene in maturity-onset diabetes of the young (MODY3). *Nature*, **384**, 455-458, 1996b.

Yamauchi T., Kamon J., Waki H., Terauchi Y., Kubota N., Hara K., Mori Y., Ide T., Murakami K., Tsuboyama-Kasaoka N., Ezaki O., Akanuma Y., Gavrilova O., Vinson C., Reitman M.I., Kagechika H., Shudo K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M., Froguel P. and Kadowaki T. The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. *Nat. Med.*, **7**, 941-946, 2001.

Yamauchi T., Kamon J., Minokoshi Y., Ito Y., Waki H., Uchida S., Yamashita S., Noda M., Kita S., Ueki K., Eto K., Akanuma Y., Froguel P., Foufelle F., Ferre P., Carling D., Kimura S., Nagai R., Kahn B.B. and Kadowaki T. Adiponectin stimulates glucose utilisation and fatty acid oxidation by activating AMP-activated protein kinase. *Nat. Med.*, **8**, 1288-1295, 2002.

Yu J., Zhang Y., McIlroy J., Rodorf-Nikolic T., Orr G.A. and Backer J.M. Regulation of the p85/p110 phosphatidylinositol 3'-kinase: Stabilisation and inhibition of the p110 α catalytic subunit by the p85 regulatory subunit. *Mol. Cell. Biol.*, **18**, 1379-1387, 1998.

Zhang Y., Proenca R., Maffei M., Barone M., Leopold L. and Friedman J.M. Positional cloning of the mouse obese gene and its human homologue. *Nature*, **372**, 425-431, 1994.

Zietz B., Herfarth H., Paul G., Ehling A., Muller-Ladner U., Scholmerich J. and Schaffler A. Adiponectin represents an independent cardiovascular risk factor predicting serum HDL-cholesterol levels in type 2 diabetes. *FEBS Lett.*, **545**, 103-104, 2003.

Zinman B., Hanley A.J.G., Harris S.R., Kwan J. and Fantus I.G. Circulating tumor necrosis factor-α concentrations in a native Canadian population with high rates of type 2 diabetes mellitus. *J. Clin. Endocrinol. Metab.*, **84**, 272-278, 1999.

Zouali H., Hani E.H., Phillipi A., Vionnet N., Beckmann J.S., Demenais F. and Froguel P. A susceptibility locus for early onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20 q proximal to the phosphoenolpyruvate carboxykinase gene. *Hum. Mol. Genet.*, **6**, 1401-1408, 1997.

Zouali H., Velho G. and Froguel P. Polymorphism of the glycogen synthase gene and non-insulin-dependent diabetes mellitus [letter]. *N. Engl. J. Med.*, **328**, 1568, 1993.

7.2 ELECTRONIC REFERENCES

NCBI, National Centre for Biotechnology Information, National Library of Medicine, National Institutes of Health, (Bethesda, MD), http://www.ncbi.nlm.nih.gov, 2004.

Das K., Lin Y., Widen E., Zhang Y. and Scherer P.E. Homo sapiens adipocyte complement-related protein (ACRP30) gene, promoter region and 5' flanking sequence. GenBank accession number AF304467, 2000.

Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M., Hinokio Y., Lindner T.H., Mashima H., Schwarz P.E.H., del Bosque-Plata L., Horikawa Y., Oda Y., Yoshiuchi I., Colilla S., Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schulz J., Baier L.J., Bogardus C., Groop L., Boerwinkle E., Hanis C.L. and Bell G.I. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. GenBank accession number AF158748, 2000.

Jackson J.L., O' Malley P., Wei G., Berbano E., Dezee K. and Shimeall W. Meta-analysis 101. http://www.sgim.org, 2004.

Liu W., Icitovic N., Shaffer M.L. and Chase G.A. The impact of population heterogeneity on risk estimation in genetic counseling. *BMC Medical Genetics*, **5**, 18-24, http://biomedcentral.com/1471-2350-5-18, 2004.

Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C., Alsbrooks S.L., Amaratunge H.C., Are J.R., Ayele M., Banks T., Barbaria J., Benton J., Bimage K., Blankenburg K., Bonnin D., Bouck J., Bowie S., Brieva M., Brown E., Brown M., Bryant N.P., Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C., Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A., He X., Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B., Homsi F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J., Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loulseged H., Lozado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J., Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E., Massey E., Mawhiney E., McLeod M.P., Meador M., Mei G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K., Moore S., Morgan M., Moorish T., Morris S., Moser M., Neal D., Nelson D., Newtson J., Newtson N., Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokenkwo S., Oguh M., Okwuonu G., Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y., Rives M., Rojas A., Rojubokan I., Rolfe M., Ruiz S., Savery G., Scherer S., Scott G., Shen H., Shooshtari N., Sisson I., Sodergren E., Sonaike T., Sparks A., Stanley H., Stone H., Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H., Tansey J., Taylor C., Taylor T., Telfrod B., Thomas N., Thomas S., Usmani K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q., Wang S., Ward-Moore S., Warren R., Washington C., Watlington S., Williams G., Williamson A., Wleczyk R., Wooden S., Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Naylor S.L., Weinstock G. and Gibbs R. Complete sequence of *Homo sapiens* 3 BAC RP11-240O9 (Roswell Park Cancer Institute Human BAC Library). GenBank accession number AF304467, 2000.

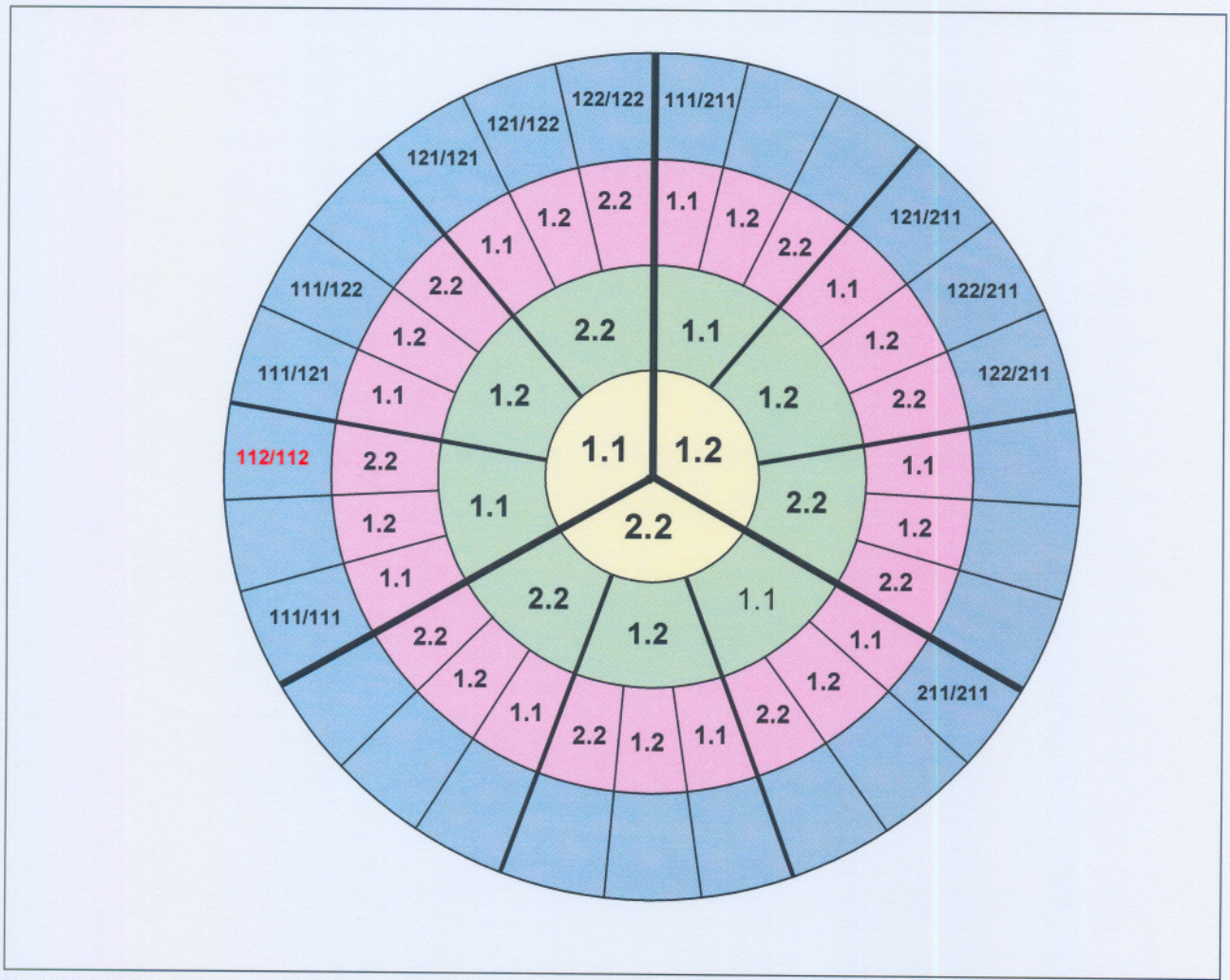
Rodgers A., Ezzati M., Van der Hoorn S., Lopez A.D., Lin R.-B., Murray C.J.L. and Comparative risk assessment collaborating group. Distribution of major health risks: findings from the global burden of disease study. PLOS Medicine, **1**, e27-e38, http://www.plosmedicine.org, 2004.

APPENDIX A

HAPLOTYPE COMBINATION ASSIGNMENT

The method for haplotype assignment within this investigation is presented in Graph A.1. This graph was generated according to the haplotype distributions present in the Mexican American, Finnish and German populations (Horikawa *et al.*, 2000). Only certain haplotypes were detected within these reported populations and therefore only specific haplotype combinations existed prior to the current study. The constituent genotypes from which these haplotypes were inferred are presented below.

Graph A.1: Graphical representation of the assignment of haplotype combinations utilised in this investigation[1]



Text presented in the yellow (□) circle represents the genotype at UCSNP-43. Text presented in the green (□) circle represents the genotype at UCSNP-56. Text indicated in the pink (□) circle represents the genotype at UCSNP-63. Text indicated in the blue (□) circle represent the haplotype combinations assigned. Text indicated in orange (xxx) highlights a haplotype combination that is present at a frequency greater than 0.05 within the black Southern African population but not at an appreciable level in the Mexican American and the German populations (Horikawa *et al.*, 2000).

[1] Adapted with permission of P. Schwarz, 2003. Originally constructed by N.J. Cox.

APPENDIX B

ACADEMIC OUTPUTS DURING THE PERIOD OF THE STUDY

Research was presented at the following national and international meetings during the period of this study. The presenting author's name is underlined in each case.

B.1 PRESENTATIONS AT INTERNATIONAL CONFERENCES

B.1.1 **53rd Annual meeting of the American Society of Human Genetics**, Los Angeles, USA., November 2003. Olckers A., **Towers G.W.**, Wessels M.N., Rheeder P. and Schwarz P. Genotypic variation of specific SNPs in the Calpain 10 gene within a cohort of the Black South African population (poster presentation).

B.1.2 **5th Annual Symposium of Deutsche Gesellschaft für Ernährung-Junge Forschung Aktiv Perspektivenkonferenz** (*English translated title: German Nutrition Society-Young Researchers Active Perspectives Conference*), Dresden, Saxony, Germany, October 2003. **Towers G.W.**, Olckers A. and Schwarz P. Haplotype combination 11.21 in the promoter of the adiponectin gene (APM-1) is associated with increased diabetes risk in the Caucasian population (platform presentation).

B.2 PRESENTATIONS AT NATIONAL CONFERENCES

B.2.1 **10th Biennial Congress of the South African Society of Human Genetics**, Durban, KwaZulu Natal, South Africa, May 2003. **Towers G.W.**, Schwarz P., Rheeder P. and Olckers A. The adiponectin SNP profile of a Black South African and a German diabetic population: a comparative study (poster presentation).

Manuscripts generated during the period of this study are listed in the following section. Distinction is made between manuscripts which were published and those submitted or under construction.

B.3 PUBLICATIONS IN INTERNATIONAL PEER REVIEWED SCIENTIFIC JOURNALS

B.3.1 Fischer S., Julius U., Hanefeld M., Fücker K., Gräßler J., **Towers G.W.**, Schwanebeck U., Schulze J. and Schwarz P. Das vorkommen vom genotyp 1.1 im SNP-44 des CAPN-10-gens ist im stadium des Typ-2-diabetes mit erhöhten insulin- und c-peptid-werten assoziiert (*Printed in German, English translated title*: Presence of allele 1 at SNP 44 in CAPN10 is associated with increased

insulin and c-peptide values in patients with Type 2 diabetes) *Diabetes und Stoffwechsel* (Diabetes and Metabolism), **13**, 3-9, 2004.

B.4 MANUSCRIPTS SUBMITTED FOR REVIEW

B.4.1 Schwarz P.E.H., **Towers W.**, Grässler J., Schwanebeck U., Rietzsch H., Fischer S., Julius U., Olckers A., Vasseur F. and Schulze J. Haplotypes in the promoter region of the *ACDC* gene (APM1) are associated with increased diabetes risk in a German Caucasian population. *Exp. Clin. Endocrinol. Diabetes*, submitted, 2004.

B.4.2 Schwarz P.E.H., **Towers G.W.**, Fuecker K., Grässler J., Fischer S., Kurktschiev T.T., Schulze J., Olckers A., Bornstein S.R., Hanefeld M. and Vasseur F. Hypoadiponectinaemia is associated with progression toward type 2 diabetes and genetic variation in the promoter of the *ACDC* gene. *Diabetologia*, submitted, 2005.

B.4.3 Schwarz P.E.H., **Towers G.W.**, Fischer S., Grässler J., Temelkova-Kurktschiev T., Olckers A., Julius U., Hanefeld M. and Schulze J. Presence of allele 1 at SNP 44 in *CAPN10* is associated with increased insulin values in newly diagnosed type 2 diabetics. *Diabetes*, submitted, 2005.

B.4.4 Olckers A., **Towers G.W.**, van der Merwe A., Schulze J. and Schwarz P.E.H. *CAPN10* haplotypes detected in high frequencies in type 2 diabetics and controls in the black South African population, in contrast to non-African populations. *Diabetologia*, submitted, 2005.

B.4.5 **Towers G.W.**, Schwarz P.E.H., Van der Merwe A., Schulze J. and Olckers A. Protective effect against type 2 diabetes identified within the *ACDC* gene, in a black South African diabetic cohort. *Diabetes*, submitted, 2005.

B.5 MANUSCRIPTS CURRENTLY UNDER CONSTRUCTION

B.5.1 Schwarz P.E.H., **Towers G.W.**, van der Merwe A., Perez-Perez L., Schulze J. and Olckers A. Meta-analysis of the C-11377G alteration in the adiponectin (ACDC) gene in the German, Cuban and Black South African population.

B.5.2 **Towers G.W.**, Perez-Perez L., Schulze J., Olckers A. and Schwarz P.E.H. CAPN10 haplotype combination 122/122 is associated with increased risk for type 2 diabetes in the Cuban population.