# Evaluation of common genetic variants associated with Type 2 Diabetes susceptibility in a black South African population

**T Chikowore**

**23917245**

Dissertation submitted in fulfilment of the requirements for the degree *Magister Scientiae* in Nutrition at the Potchefstroom Campus of the North-West University

| | |
|---|---|
| Supervisor: | Dr T van Zyl |
| Co-supervisor: | Dr KR Conradie |
| Asistant Supervisor: | Dr GW Towers |

September 2014

It all starts here ™

NORTH-WEST UNIVERSITY ®
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

# Abstract

**Introduction:** The continual increase of type 2 diabetes (T2D) prevalence is a global public health concern. The aetiology of T2D has not been fully elucidated and this is hampering the development of effective preventative and curative interventions to curb the T2D burden. Although much has been done to elucidate the environmental risk factors associated with T2D, little is known about the precise genetic risk factors that predispose people to it. There is limited knowledge about the common variants associated with T2D risk in the black South African population. However, evidence of shared common variants associated with T2D among people of different ethnicities has been documented. Nonetheless, the majority of the common variants that have been reported to be associated with T2D in other ethnicities are still yet to be evaluated in the black South African population.

**Objectives:** The aim of this study was to evaluate the association of previously reported common genetic variants with T2D susceptibility, as indicated by impaired glucose tolerance (IGT), in a black South African population of Tswana descent.

**Methods:** This study was a case-control study of 180 cases and 180 controls nested in the Prospective Urban Rural Epidemiology (PURE) study baseline data, which was collected in 2005. The DNA samples of the participants were genotyped for 77 single nucleotide polymorphisms (SNPs), using Illumina® VeraCode technology on the BeadXpress® platform. The gPlink software was used to evaluate the standard genetic models of disease penetrance for the association of the common variants with impaired glucose tolerance (IGT) while adjusting for age, sex and body mass index.

**Results:** Four out of the 66 SNPs that were evaluated through the genetic association tests in this study were noted to be significantly associated with IGT ($p < 0.05$). Of the four SNPs, only rs1436955 was associated with an increase in T2D risk, while the other three variants, rs831571, rs8050136 and rs7542900, were noted to be associated with a decreased risk of T2D. However, none of the four SNPs was significantly associated with IGT after correcting for multiple testing ($p < 0.05$).

**Conclusions:** Black South Africans of Tswana descent might not share common variants associated with T2D risk, as indicated by IGT in other ethnicities. Well-powered studies are required to evaluate the association of common variants with T2D risk in this population group. The results from this study emphasise the need for

population-specific variants to assess the genetic susceptibility of complex diseases such as T2D in the black South African population.

**Key words:** Type 2 diabetes; genetics; SNPs; common genetic variants; GWAS; South African black population.

# Acknowledgements

I would like to thank the Lord for uniting me with the following people who contributed immensely to the successful completion of this dissertation:

I am grateful to my supervisors, **Dr Tertia van Zyl** and **Dr Karin Conradie,** for their belief in me and particularly the motivation they continually rendered to inspire me to complete this work. I am grateful for their guidance and commitment to my work, even when they were inundated with pressing social and academic responsibilities. May God reward them for their kindness.

I would like also to thank **Dr Wayne Towers** for his cheerful spirit. Those lighter moments helped relieve the stress of completing this task. Thank you for the guidance, motivation, critical analysis and assistance with the editing of this dissertation.

I am grateful to **Dr Suria Ellis** for assistance with the statistical analysis. I would also like to acknowledge **Dr Zané Lombard** and **Liesl Hendry** for teaching me how to use gPlink software, which enabled me to analyse the genetic association tests with ease. I would also like to thank **Herman Myburgh** for assistance with the laboratory work. Much thanks also goes to **Barbara Bradley** for the language editing.

Special thanks go to my wife, **Priscilla Masawi,** for her continual motivation and assurance that I could complete this work. I am also grateful to the **South African Sugar Association** (SASA) for their funding, which helped us to conduct this research. I would like to acknowledge the **PURE research team**, **funders** and **participants**. I would like to acknowledge Prof Antonel Olckers, the PRIMER team, DNAbiotec Pty Ltd and the Centre for Genome Reseach, NWU for the financial contribution and carrying out of the oral glucose tolerance tests. Above all, I am greatly indebted to the **Lord** for the power to complete this work and HIS love, which saw me through it all. To HIM be the glory for ever and ever for fulfilling my dreams and being continually faithful to an undeserving person like me. AMEN

# Table of Contents

**Abstract** ..... **i**
**Acknowledgements** ..... **iii**
**List of Tables** ..... **vii**
**List of Figures** ..... **viii**
**List of Abbreviations** ..... **ix**
**List of symbols and units** ..... **xii**
**List of equations** ..... **xiii**

**CHAPTER 1 INTRODUCTION** ..... **1**
1.1 Problem Statement ..... 3
1.2 Aim ..... 4
1.3 Specific Objectives ..... 4
1.4 Structure of the dissertation ..... 4
1.5 Contribution of the researchers in this study ..... 5

**CHAPTER 2 LITERATURE REVIEW** ..... **6**
2.1 Diabetes, symptoms and complications ..... 6
2.2 Classification of Diabetes Mellitus ..... 7
2.2.1 Impaired glucose tolerance ..... 8
2.2.2 Type 1 diabetes ..... 9
2.2.3 Type 2 diabetes ..... 9
2.2.4 Other forms of diabetes ..... 10
2.3 Methods for Type 2 diabetes diagnosis ..... 11
2.3.1 Fasting plasma glucose determination test ..... 12
2.3.2 Oral glucose tolerance test determination ..... 13
2.3.3 Glycated haemoglobin determination ..... 13
2.4 Framework for the Aetiology of Type 2 diabetes ..... 14
2.4.1 Gene-environmental interactions in infancy ..... 15
2.4.2 Diet, physical activity and Type 2 diabetes susceptibility ..... 16
2.4.3 Microbiome and Type 2 diabetes susceptibility ..... 17
2.4.4 Endocrine-disrupting chemicals and Type 2 diabetes susceptibility ..... 17
2.5 Heritability of Type 2 diabetes ..... 18
2.5.1 Linkage analysis ..... 19
2.5.2 Candidate gene association studies for type 2 diabetes ..... 19
2.5.3 Genome-wide association studies and type 2 diabetes genetic susceptibility ..... 20
2.5.4 Genome-wide association studies of type 2 diabetes in European populations ..... 22
2.5.5 Genome-wide association studies of type 2 diabetes in other populations ..... 24
2.6 Missing Heritability of GWAS-identified Type 2 diabetes associated SNPs ..... 29
2.6.1 Rare gene variants and missing Type 2 diabetes heritability ..... 29
2.6.2 Copy number variations and missing Type 2 diabetes heritability ..... 30
2.6.3 Gene-gene interactions and missing heritability ..... 30
2.6.4 Gene-environment interactions and missing heritability ..... 31
2.6.5 Epigenetics and missing heritability of Type 2 diabetes ..... 32
2.6.6 Systems biology and missing Type 2 diabetes heritability ..... 32
2.7 Type 2 diabetes genetic studies in Africa ..... 33

2.7.1 Shared type 2 diabetes gene variants among different ethnicities ..... 34
2.7.2 GWAS-determined common variants among black South Africans and T2D genetic susceptibility ..... 35
2.8 Summary ..... 35

**CHAPTER 3 METHODOLOGY ..... 37**
3.1 Study Design ..... 37
3.2 Study Population ..... 38
3.3 Sample Size calculation ..... 39
3.4 Oral glucose tolerance test ..... 39
3.5 Anthropometry ..... 39
3.6 Questionnaires ..... 40
3.7 Genetic Analysis ..... 41
3.7.1 Literature search of SNPs associated with Type 2 diabetes ..... 41
3.7.2 Pre-assessment of SNPs for genotyping ..... 42
3.7.3 DNA isolation ..... 42
3.7.4 Normalisation of DNA ..... 43
3.7.5 Genotyping with Illumina BeadXpress® Technology ..... 43
3.7.6 Validation of the BeadXpress® genotyping ..... 44
3.7.7 Primer design ..... 45
3.7.8 Amplification conditions ..... 46
3.7.9 Sequencing ..... 46
3.8 Statistical Analysis ..... 48
3.8.1 Descriptive statistics ..... 48
3.8.2 Calculation of the distribution patterns of the selected SNPs ..... 48
3.8.3 Testing for genetic association of SNPs with IGT ..... 49
3.8.4 Analysis strategy for the genetic association tests ..... 50

**CHAPTER 4 RESULTS AND DISCUSSION ..... 53**
4.1 Descriptive statistics of research participants ..... 53
4.2 DNA Isolation ..... 56
4.3 Analysis of the BeadXpress® results for selected snps ..... 56
4.3.1 Pre-assessment of SNPs for BeadXpress® genotyping ..... 57
4.3.2 Analysing data from the BeadXpress® reader ..... 60
4.3.3 Evaluation of the BeadXpress® Reader genotyping controls ..... 61
4.3.3.1 Allele-specific extension controls ..... 62
4.3.3.2 Contamination controls ..... 63
4.3.3.3 Polymerase chain reaction uniformity controls ..... 64
4.3.3.4 Extension gap controls ..... 65
4.3.3.5 First hybridisation controls ..... 66
4.3.3.6 Second hybridisation controls ..... 67
4.3.4 Evaluation of the cluster plots of the selected SNPs ..... 67
4.3.5 Validation of the BeadXpress® genotyping ..... 70
4.4 Hardy-weinberg equilibrium status of the selected common variants ..... 71
4.5 Genetic association analysis ..... 75
4.5.1 SNP rs831571 ..... 77
4.5.2 SNP rs8050136 ..... 79
4.5.3 SNP rs7542900 ..... 81
4.5.4 SNP rs1436955 ..... 82

4.5.5 Possible reasons for discrepancy of association patterns of selected common variants ..... 83
4.5.5.1 Low power of study ..... 83
4.5.5.2 Genetic diversity of black South Africans compared to other ethnicities ..... 84
4.5.5.3 Intra-genetic diversity ..... 85
4.6 Summary of the results ..... 85

**CHAPTER 5 CONCLUSION ..... 87**
5.1 Summary of the main findings ..... 87
5.2 Implications of this study ..... 88
5.3 Limitations ..... 89
5.4 Recommendations ..... 89
5.5 Conclusion ..... 90

**REFERENCES ..... 92**

**APPENDIX A**
**Additive genetic model univariate and multivariate results for the selected SNPs ..... 119**

**APPENDIX B**
**General genotypic genetic model univariate results for the selected SNPs ..... 122**

**APPENDIX C**
**Dominant genetic model univariate and multivariate results for the selected SNPs ..... 125**

**APPENDIX D**
**Recessive genetic model univariate for the selected SNPs ..... 128**

**APPENDIX E**
**CATT genetic model univariate results for the selected SNPs ..... 131**

# List of Tables


| | | |
|---|---|---|
| Table 1.1 | Contributions of the researchers in this study | 5 |
| Table 2.1 | Clinical stages of diabetes | 8 |
| Table 2.2 | Aetiological types of diabetes | 11 |
| Table 2.4 | The list of SNPs that have been associated with T2D in various ethnicities through GWAS | 26 |
| Table 4.1 | Descriptive statistics of the research participants | 56 |
| Table 4.2 | Results of the pre-evaluation of the SNPs for BeadXpress® genotyping | 58 |
| Table 4.3 | IllumiCode sequence IDs of controls and their expected outcomes | 61 |
| Table 4.4 | Comparisons of the BeadXpress and Sanger Sequencing genotyping results for selected samples | 70 |
| Table 4.5 | HWE status of the selected common variants | 73 |
| Table 4.7 | Multivariate genetic association results for the selected SNPs | 77 |
| Table A.1 | Univariate test results for the additive genetic model | 119 |
| Table B.1 | Univariate test results of the general genotypic genetic model | 122 |
| Table C.1 | Univariate test results for the dominant genetic model | 125 |
| Table D.1 | Univariate test results of the recessive genetic model | 128 |
| Table E.1 | Univariate test results of the CATT genetic model | 131 |

# List of Figures


Figure 2.1 Interaction of genetic and environmental factors in the development of T2D. ..... 15
Figure 2.2 GWAS case control study design ..... 22
Figure 2.3 Spectrum of disease-causing variants and their effect sizes. ..... 30
Figure 2.4 The future of systems biology approach for studying T2D ..... 33
Figure 3.1 Genetic models of disease penetrance ..... 50
Figure 3.2 Analysis strategy ..... 52
Figure4.1 Scatter plot for the 2 hour glucose concentrations of the study participants……..54
Figure 4.2 Allele-specific extension controls ..... 63
Figure 4.3 Contamination controls..... 64
Figure 4.4 Polymerase chain reaction uniformity controls ..... 65
Figure 4.5 Extension gap controls ..... 66
Figure 4.6 First hybridisation controls..... 66
Figure 4.7 Second hybridisation controls ..... 67
Figure 4.8 Genotype cluster plot for the SNP rs8050136 ..... 69
Figure 4.9 Genotype cluster plot for the SNP rs17797882 ..... 69
Figure 4.10 Electropherogram for the SNP rs896854 ..... 71

# List of Abbreviations

| | |
|---|---|
| *ACHE* | Acetylcholinesterase (Yt blood group) |
| ADA | American Diabetes Association |
| *ADAM30* | ADAM metallopeptidase domain 30 |
| *ADAMTS9* | ADAM metallopeptidase with thrombospondin type 1 motif, 9 |
| *ADCY5* | Adenylate cyclase 5 |
| ADT | Assay design tool |
| AGEs | Advanced glycation end products |
| *AP3S2* | Adaptor-related protein complex 3, sigma 2 subunit |
| *ARAP1* | ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1 |
| ASO | Allele-specific oligonucleotides |
| *BARX2* | BARX homeobox 2 |
| *BCL11A* | B-cell CLL/lymphoma 11A (zinc finger protein) |
| BPA | Bisphenol |
| CATT | Cochran-Armitage test for trend |
| *C14orf70* | Long intergenic non-protein coding RNA 523 |
| *C2CD4A* | C2 calcium-dependent domain containing 4A |
| *C2CD4B* | C2 calcium-dependent domain containing 4B |
| *C6orf57* | Chromosome 6 open reading frame 57 |
| *CAMK1D* | Calcium/calmodulin-dependent protein kinase ID |
| *CAPN10* | Calpain 10 |
| CD/CV | Common disease/common variant hypothesis |
| *CDC123* | Cell division cycle 123 |
| *CDKAL1* | CDK5 regulatory subunit associated protein 1-like 1 |
| *CDKN2A* | Cyclin-dependent kinase inhibitor 2A |
| *CDKN2B* | Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4) |
| *CETN3* | Centrin, EF-hand protein, 3 |
| *CHCHD9* | Coiled-coil-helix-coiled-coil-helix domain containing 2 pseudogene 9 |
| *CMIP* | c-Maf inducing protein |
| CNV | Copy number variations |
| *CR2* | Complement component (3d/Epstein Barr virus) receptor 2 |
| DCCT | Diabetes Control and Complications Trial |
| *DCD* | Dermcidin |
| *DGKB-TMEM195* | Diacylglycerol kinase, beta 90kDa-alkylglycerol monooxygenase |
| DIAGRAM | Diabetes Genetics Replication and Meta-analysis |
| DNA | Deoxyribonucleic acid |
| *DUSP9* | Dual specificity phosphatase 9 |
| EDCs | Endocrine disrupting chemicals |
| *F3* | Coagulation factor III (thromboplastin, tissue factor) |
| *FITM2-R3HDML-HNF4A* | Fat storage-inducing transmembrane protein 2-R3H domain containing-like-hepatocyte nuclear factor 4, alpha |
| FPG | Fasting plasma glucose |
| *FTO* | Fat mass and obesity-associated |
| *GALNTL4* | Polypeptide N-acetylgalactosaminyltransferase 18 |

| | |
|---|---|
| *GCC1-PAX4* | GRIP and coiled-coil domain containing 1- paired box 4 |
| *GCK* | Glucokinase, (hexokinase 4) |
| *GCKR* | Glucokinase (hexokinase 4) regulator |
| *GLIS3* | GLIS family zinc finger |
| *GLIS3* | GLIS family zinc finger 3 |
| *GRB14* | Growth factor receptor-bound protein 14 |
| GWAS | Genome-wide association studies |
| H3 Africa | Human, Heredity and Health in Africa |
| HbA1C | Glycated haemoglobin |
| *HHEX* | Haematopoietically expressed homeobox |
| *HLA* | Human leukocyte antigen |
| *HMG20A* | High mobility group 20A |
| *HMGA2* | High mobility group AT-hook 2 |
| *HNF1A* | HNF1 homeobox A |
| *HNF1B* | HNF1 homeobox B |
| *HNF4A* | Hepatocyte nuclear factor 4, alpha |
| *HUNK* | Hormonally up-regulated Neu-associated kinase |
| HWE | Hardy-Weinberg equilibrium |
| IFG | Impaired fasting glucose |
| *IGF2BP2* | Insulin-like growth factor 2 mRNA binding protein 2 |
| IGT | Impaired glucose tolerance |
| *IRS1* | Insulin receptor substrate 1 |
| *ITGB6* | Integrin, beta 6 |
| *JAZF1* | JAZF zinc finger 1 |
| *KCNJ11* | Potassium inwardly-rectifying channel, subfamily J, member 11 |
| *KCNK16* | Potassium channel, subfamily K, member 16 |
| *KCNQ1* | Potassium voltage-gated channel, KQT-like subfamily, member 1 |
| *KIF11* | Kinesin family member 11 |
| *KLF14* | Kruppel-like factor 14 |
| LD | Linkage disequilibrium |
| *LGR5* | Leucine-rich repeat containing G protein-coupled receptor 5 |
| *LPIN2* | Lipin 2 |
| LSO | Locus-specific oligonucleotides |
| *MAEA* | Macrophage erythroblast attacher |
| MAGIC | Meta-analysis of Glucose and Insulin-related Traits Consortium |
| MODY | Maturity onset diabetes of the young |
| *MTNR1B* | Melatonin receptor 1B |
| NCDs | Non-communicable diseases |
| NDDG | National Diabetes Data Group |
| *NOTCH2* | Notch 2 |
| OGTT | Oral glucose tolerance test |
| *PCNXL2* | Pecanex-like 2 (Drosophila) |
| PCR | Polymerase chain reaction |
| *PEPD* | Peptidase D |
| *PEX5L* | Peroxisomal biogenesis factor 5-like |

| | |
|---|---|
| *PLS1* | Plastin 1 |
| *PPARG* | Peroxisome proliferator-activated receptor gamma |
| *PRC1* | Protein regulator of cytokinesis 1 |
| *PROX1* | Prospero homeobox 1 |
| *PSMD6* | Proteasome (prosome, macropain) 26S subunit, non-ATPase, 6 |
| *PTPRD* | Protein tyrosine phosphatase, receptor type, D |
| PURE | Prospective Urban Rural Epidemiology study |
| *RBM43* | RNA binding motif protein 43 |
| *RBMS1* | RNA binding motif, single stranded interacting protein 1 |
| *RND3* | Rho family GTPase 3 |
| *SLC30A8* | Solute carrier family 30 (zinc transporter), member 8 |
| *SLC44A3* | Solute carrier family 44, member 3 |
| SNPs | Single nucleotide polymorphisms |
| *SPRY2* | Sprouty homolog 2 (Drosophila) |
| *SRR* | Serine racemase |
| *ST6GAL1* | ST6 beta-galactosamide alpha-2,6-sialyltranferase 1 |
| *SYN2* | Synapsin II |
| T1D | Type 1 diabetes |
| T2D | Type 2 diabetes |
| *TCERG1L* | Transcription elongation regulator 1-like |
| *TCF7L2* | Transcription factor 7-like 2 |
| *THADA* | Thyroid adenoma associated |
| *TLE4* | Transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila) |
| *TMEM45B* | Transmembrane protein 45B |
| *TP531NP1* | Tumour protein p53 inducible nuclear protein 1 |
| *TSPAN8* | Tetraspanin 8 |
| *UBE2E2* | Ubiquitin-conjugating enzyme E2E 2 |
| *VEGFA* | Vascular endothelial growth factor A |
| *VPS26A* | Vacuolar protein sorting 26 homolog A (S. pombe) |
| *WFS1* | Wolfram syndrome 1 |
| WHO | World Health Organisation |
| *WWOX* | WW domain containing oxidoreductase |
| *ZBED3* | Zinc finger, BED-type containing 3 |
| *ZFAND3* | Zinc finger, AN1-type domain 3 |
| *ZFAND6* | Zinc finger, AN1-type domain 6 |
| *ZPLD1* | Zona pellucida-like domain containing 1 |

# List of symbols and units

| | |
|---|---|
| β | beta |
| $X^2$ | Chi square |
| Δ | delta |
| °C | degree centigrade |
| = | equal |
| g | gravitational force |
| g | gram |
| > | greater than |
| ≥ | greater than or equal |
| $h^2$ | heritability |
| L | litre |
| kg | kilogram |
| kJ | kilojoules |
| kg/m2 | kilogram per metre squared, unit of body mass index |
| < | less than |
| ≤ | less than or equal to |
| µ | micro |
| µL | microlitre |
| m | milli |
| mL | millilitre |
| $mmol.L^{-1}$ | millimoll per litre |
| mol | mole |
| M | molecular weight |
| x | multiply |
| - | negative minus |
| n | number of subjects |
| p | p-value, indicates statistical significance |
| % | percentage |
| ± | plus minus |
| $\lambda_s$ | sibling relative risk |

# List of equations

Equation 3.1 Formula for calculating BMI…… 40
Equation 3.2 Normalisation of DNA…… 43

# CHAPTER 1

## INTRODUCTION

The increasing global prevalence of type 2 diabetes mellitus (T2D) is a great cause for concern. The International Diabetes Federation estimates that the number of people with T2D will increase from 366 million in 2011 to 552 million in 2030 (Sanghera & Blackett, 2012). Although this is a global prediction, the black South African population has not been spared from this increasing prevalence of T2D. The highest age-standardised prevalence of T2D in Sub-Saharan Africa of 13.1% was reported among black South Africans in Cape Town (Peer *et al*., 2012). T2D is posing a threat to the economic development of many countries, as it is associated with increased mortality in the productive age category of South Africa (Bradshaw *et al*., 2007). According to Bradshaw *et al.* (2007), T2D is associated with increased mortality of 14% due to ischaemic heart disease, 10% due to stroke, 12% due to hypertension and 12% due to renal disease in the productive age category above 30 years in South Africa.

Evidence exists that 80% of the mortality due to non-communicable diseases (NCDs), such as T2D, can be prevented through diet and exercise interventions (WHO, 2005). Alarmingly, because of the asymptomatic nature of the actual onset of T2D, it is estimated that 30-80% of type 2 diabetic patients remain undiagnosed and upon diagnosis 20% of them would have already developed complications (Saudek *et al.,* 2008; Amod *et al.,* 2012). In a number of studies, T2D-related complications were detected in individuals who had impaired glucose tolerance (IGT) and had not yet developed T2D (Nguyen *et al.,* 2007; Sumner *et al.,* 2003; Barr *et al.,* 2007). A four-to-seven-year lag is thought to exist between the actual onset of T2D and its diagnosis (Harris, 1993). The Society for Endocrinology, Metabolism and Diabetes of South Africa, in its recent guidelines for the management of T2D, urged the need for developing early T2D screening tools for high-risk groups, coupled with timely interventions as the prime approach to curbing this growing pandemic (Amod *et al.,* 2012). However, the development of timely interventions to curb the increase in T2D is being hampered by limited understanding of the precise aetiology of T2D.

The aetiology of T2D is postulated to involve the complex interaction of genetic and environmental factors (Goh, 2012). According to a review of research on T2D in Sub-Saharan Africa, a number of studies have been done to assess the effects of environmental factors on T2D in African settings (Mbanya *et al*., 2010). However, more studies are required to clarify the manner in which genetic susceptibility modifies the effect of the environmental risk factors for T2D. Urbanisation and the adoption of Western dietary patterns are the major environmental risk factors that have been associated with the increase in T2D prevalence worldwide (Qi *et al*., 2013). Although evidence now exists that T2D risk due to the adoption of Western dietary lifestyles varies with genetic risk (Qi *et al*., 2013), knowledge about the precise genetic factors that constitute T2D genetic susceptibility is still limited.

T2D is hypothesised to have a strong genetic aetiology (Florez, 2013). A high concordance rate of 96% for developing T2D was reported among monozygotic twins (Lo *et al.,* 1991*)*. In addition, 40% of first-degree relatives of T2D patients were reported to develop T2D, compared to a baseline population risk of 6%, suggesting a strong underlying genetic component (Kobberling, 1982). However, the twin studies of T2D genetic susceptibility are unable to provide evidence of the precise genetic factors underlying the heritability of this disease.

The search for genetic variants associated with T2D continues. Ninety-six percent of the genome-wide association studies (GWASs), which contributed much of the knowledge regarding the genetic variants currently associated with T2D, were undertaken in Caucasian populations (Rosenberg *et al.,* 2010). In view of the genetic diversity that exists between people in the world, the GWAS findings in Caucasian individuals have been deemed unsuitable for determining disease risk in other populations (Sanghera & Blackett, 2012). The Human, Heredity and Health in Africa (H3 Africa) project is an example of a recent initiative that has been developed to research the genetic aetiology of NCDs such as T2D in the African population (Ramsay, 2012). Therefore more T2D genetics studies are required in the African setting, since the GWAS findings from the European populations cannot be extrapolated to determine genetic susceptibility in other populations.

To date 79 genetic variants have been associated with T2D among the European, Asian and Mexican populations (Sanghera & Blackett, 2012). However, the currently reported variants can only account for about 10% of the observed heritability of T2D (Sanghera & Blackett, 2012). Even though the current genetic variability only explains 10% of the heritability of T2D in non-Africans, there is a need to assess it among African populations as well. There is now growing evidence of T2D genetic variants that are shared among people of different ethnicities (Waters *et al*., 2010). The Europeans and North African Arabs were noted to have similar disease association patterns for 13 of the genetic variants associated with T2D (Cauchi *et al.,* 2012). European Americans, African Americans, Latino Americans, Japanese Americans and Native Hawaiians have also been reported to share 14 genetic variants associated with type T2D (Waters *et al*., 2010). However, only one single genetic variant in the *TCF7L2* gene out of five other loci associated with T2D in Europeans was noted to be associated with T2D in black South Africans of Zulu descent (Pirie *et al.,* 2010). It is therefore plausible that the black South African population may also share more than one T2D associated genetic variant with other ethnicities, as most of the currently known 79 variants associated with T2D have not been evaluated in this population group.

## 1.1 PROBLEM STATEMENT

The strong genetic component of T2D, which is indicated by the high concordance rates in twin studies, is yet to be fully elucidated. Information on the common variants associated with T2D in the black South African population is scarce. GWAS studies of T2D in other ethnicities cannot be extrapolated to determine genetic risk in other ethnicities (Sanghera & Blackett, 2012). However, growing evidence of shared common variants associated with T2D among different ethnicities is a motivation to evaluate the reported variants in the black South African population (Cauchi *et al.,* 2012). About 79 genetic variants have been reported to be associated with T2D among the European, Asian and Mexican populations (Sanghera & Blackett, 2012). However, only a limited number of the currently reported variants associated with T2D have been evaluated in the black South African population.

This study seeks to evaluate the currently reported variants associated with T2D in other ethnicities in the black South African population of Tswana descent.

## 1.2 AIM

- To assess the association of specific single nucleotide polymorphisms (SNPs) with IGT in a black South African population.

## 1.3 SPECIFIC OBJECTIVES

**a.** To determine the type 2 diabetes risk factors (body mass index [BMI], dietary fat intake, dietary carbohydrate intake) profile of the cases and controls

**b.** To determine the allelic and genotypic distributions of the selected SNPs among individuals with IGT and IGT-negative control individuals in a black South African population.

**c.** To test the association of the selected SNPs using several genetic models

## 1.4 STRUCTURE OF THE DISSERTATION

Chapter 1 involves the elaboration of the burden of T2D globally, the factors hypothesised to be involved in the aetiology of T2D and the importance of evaluating the genetic factors associated with T2D in the black South African population. Chapter 2 is a detailed review of the literature about what is currently known about the genetics of T2D and how this study could potentially add to the knowledge about the common variants associated with T2D in black South Africans. In Chapter 3 the methods used to evaluate the common variants associated with T2D risk among black South Africans, as indicated by IGT, are described. Chapter 4 of this dissertation focuses on the results of this study and discusses their implications. Chapter 5 is the final chapter and it provides a summary of the main findings, implications of the results, the limitations of the methodology that was used and recommendations for future studies.

## 1.5 CONTRIBUTION OF THE RESEARCHERS IN THIS STUDY

The contributions of the researchers who were involved in this study are listed in Table 1.1 below.

**Table 1.1 Contributions of the researchers in this study**

| Name | Affiliation | Role in the study |
|---|---|---|
| Mr T Chikowore | Centre of Excellence in Nutrition, North-West University | MSc candidate: - Laboratory preparation of samples, data analysis, interpretation of results and writing up of the dissertation. |
| Dr T van Zyl | Centre of Excellence in Nutrition, North-West University | Supervisor:- Evaluation of laboratory work, data analysis and writing up done by student |
| Dr KR Conradie | Centre of Excellence in Nutrition, North-West University | Co-supervisor:- Evaluation of laboratory work, data analysis and writing up done by student |
| Dr W Towers | Centre of Excellence in Nutrition, North-West University | Co-supervisor:- Evaluation of data analysis, dissertation editing and writing up done by student |

# CHAPTER 2

# LITERATURE REVIEW

There is an urgent need to find ways of curbing the growing pandemic of T2D. Lack of clarity on the precise mechanism for the development of T2D is hampering current efforts to reduce the burden of this disease (Goh, 2012). Deciphering genetic susceptibility to T2D is essential to understanding the aetiology of this disease and will thus aid in the development of effective prevention and treatment interventions. This chapter is a detailed literature review of the issues pertaining to the evaluation of common genetic variants associated with T2D in the black South African population. The definition of diabetes, its signs, symptoms, classifications and the reasons for focusing on T2D are elaborated on in Section 2.1 and Section 2.2. Section 2.3 describes the methods for T2D diagnosis. The framework for the aetiology of T2D is explained in Section 2.4. Section 2.5 details the current issues and concerns regarding genetic susceptibly to T2D. The missing heritability of T2D and potential ways of accounting for it is detailed in Section 2.6. Section 2.7 provides a detailed description of the T2D research initiatives in Africa and the gaps in knowledge regarding the common variants associated with T2D risk in black South Africans. This chapter is then concluded by a summary in Section 2.8.

## 2.1 DIABETES, SYMPTOMS AND COMPLICATIONS

Diabetes mellitus is a metabolic disorder characterised by hyperglycaemia, which occurs because of either inadequate insulin secretion, insulin action, or both (Alberti & Zimmet, 1998). Thus, abnormalities in carbohydrate, fat and protein metabolism occur in diabetic individuals because of the inadequate action of insulin on target tissues (ADA, 2011). Complications such as retinopathy, nephropathy and neuropathy are long-term effects of diabetes. Diabetic individuals are also at increased risk of developing cardiovascular diseases and tuberculosis (Alberti & Zimmet, 1998).

Thirst, polyuria, blurred vision, weight loss and ketoacidosis are the common signs and symptoms of diabetes (WHO, 2011). Apart from ketoacidosis, most symptoms of diabetes

may appear less severe or absent without the use of routine biochemical assessments (Alberti & Zimmet, 1998). T2D may progress undetected for a long time, as it takes time for easily noticeable signs and symptoms to appear (Alberti & Zimmet, 1998). It is estimated that 30-80% of diabetes cases remain undiagnosed and among those diagnosed, 25% would have developed microvascular complications by the time of diagnosis (Harris, 1993; WHO, 2011).

## 2.2 CLASSIFICATION OF DIABETES MELLITUS

Since 1965, several guidelines have been developed to classify diabetes (Amod *et al*., 2012). However, in 1995 an expert committee was set up by the American Diabetes Association (ADA) to revise the guidelines for classifying diabetes in the light of current evidence and to change it from being based on pharmacological treatment to the aetiology of the disease (Alberti & Zimmet, 1998). The expert committee guidelines were published in 1999 and adopted by the World Health Organisation (WHO) for classifying diabetes (Amod *et al*., 2012). Diabetes is classified according to the aetiological types and clinical stages indicated in Table 2.1 (Amod *et al*., 2012). Regardless of the aetiological factors, the types of diabetes have similar clinical stages (Alberti & Zimmet, 1998). The clinical stages change from normoglycaemia to an intermediate phenotype, which comprises of impaired fasting glucose (IFG) or IGT and then diabetes (Alberti & Zimmet, 1998). The aetiological types of diabetes include type 1, type 2, gestational diabetes and other specific types of diabetes as indicated in Table 2.1. These aetiological classes of diabetes are elaborated on in detail in subsequent sections.

**Table 2.1 Clinical stages of diabetes**

| Stages / Types | Normoglycaemia | Hyperglycaemia | | | |
|---|---|---|---|---|---|
| | Normal glucose regulation | IGT or IFG | Diabetes Mellitus | | |
| | | | Non insulin-Requiring | Insulin-requiring for control | Insulin-requiring for survival |
| **Type 1 diabetes** | → | → | → | → | → |
| **Type 2 diabetes** | → | → | → | → | |
| **Other** | → | → | → | → | |
| **Gestational Diabetes** | → | → | → | | |

Adapted from Alberti & Zimmet. (1998). IGT = impaired glucose tolerance; IFG = impaired fasting glucose

**2.2.1 Impaired glucose tolerance** In 1979, the term IGT was introduced as an intermediate category between normal glucose tolerance and diabetes by the National Diabetes Data Group (NDDG) and WHO (NDDG, 1979; WHO, 1980). IGT is a clinical stage in the natural history of the disease but not a form of diabetes (George & Alberti, 1998). Prior to the introduction of IGT, terms such as 'borderline' diabetes and 'chemical' diabetes were used, which depicted the milder forms of diabetes (Alberti, 1996). However, the terms 'borderline' diabetes and 'chemical' diabetes did not have precise blood glucose level cut-off points (George & Alberti, 1998). The cut-off points of IGT were later derived from determining glucose levels indicative of increased risk of developing diabetes after standardised 50 g and 75 g glucose challenge tests (Jarret & Keen, 1976; Jarret & Sayegh, 1978). The ADA and a WHO consultation group reviewed the diagnostic criteria for IGT in 1998 (WHO, 1998), after which the cut-off limits were set as a two-hour post-load glucose value range of 6.7-9.9 $mmolL^{-1}$ or 120-179 $mgdL^{-1}$ for whole blood, and 7.8-11.0 $mmolL^{-1}$ or 140-199 $mgdL^{-1}$ for plasma after an oral glucose tolerance test (OGTT) had been conducted (George & Alberti, 1998). The pathophysiology of IGT is characterised by both insulin resistance and hyperinsulinaemia (Reaven *et al.,* 1989). Insulin resistance is hypothesised to cause mildly elevated glucose levels, which lead to *β*-cell glucotoxicity and IGT (Davies *et al*, 1993). Lifestyle changes, poor foetal nutrition and programming are also postulated to cause insulin resistance (Phipps *et al*., 1993). It is estimated that 2-5% of individuals with IGT develop T2D each year (Alberti, 1996). Over a period of 10 years it is

estimated that 50% of individuals with IGT would have progressed to T2D (Saad *et al*., 1988). Thus people with IGT have an increased risk of developing T2D, compared to the general population of people who are not affected by IGT.

### 2.2.2 Type 1 diabetes

The autoimmune destruction of *β-cells* in the pancreas leads to the development of type 1 diabetes (T1D) (Alberti & Zimmet, 1998). Children are mostly affected by T1D (Alberti & Zimmet, 1998). However, the term latent autoimmune diabetes in adults is used to describe T1D in adults (Humphrey *et al*., 1998). T1D has been reported to account for 5% of diabetes cases (Alberti & Zimmet, 1998). Early signs of T1D include ketoacidosis, modest fasting hyperglycaemia in the presence of infection or stress (Alberti & Zimmet, 1998). Patients with T1D become dependent on insulin for survival and are at risk of ketoacidosis (Alberti & Zimmet, 1998). Eighty-five to 90% of T1D patients have autoantibodies towards glutamic acid decarboxylase ($GAD_{65}$), insulin and islets of the pancreas, which are markers of immune destruction (Verge *et al*., 1996). Unlike T2D, patients with T1D are usually not obese (Alberti & Zimmet, 1998). Other autoimmune disorders, which include Graves's disease, Addison's disease and Hashimoto's thyroiditis, also often co-exist within T1D patients (Betterle *et al.,* 1983). Idiopathic forms of T1D have no reported aetiology, but are characterised by permanent insulinopenia and ketoacidosis without evidence of autoimmunity (Alberti & Zimmet, 1998). The interaction of the human leukocyte antigen (HLA) class II genes and environmental triggers such as cow's milk supplementation during infancy and enteroviruses are postulated to be underlying risk factors for the development of T1D (Knip *et al.,* 2005).

### 2.2.3 Type 2 diabetes

T2D is characterised by relative insulin insufficiency and resistance to its action (Harris, 1993). However, unlike T1D, insulin is not initially required for survival (Harris, 1993). Usually, T2D exists for many years without being diagnosed because hyperglycaemia will not be severe enough to cause noticeable symptoms of diabetes (Harris, 1993). Notwithstanding this, patients with T2D will be at risk of developing micro-vascular and

macro-vascular symptoms (Alberti & Zimmet, 1998). Most T2D patients are obese and it is this characteristic that is postulated to cause insulin resistance, which is the major pathological factor in the development of this disorder (Alberti & Zimmet, 1998). Insulin resistance occurs in the skeletal muscles and it is characterised by reduced responsiveness of target tissues to insulin, resulting in elevated glucose levels in the blood (Sesti, 2006). Insulin secretion among T2D patients is inadequate to adjust for the insulin resistance (Campbell & Carlson, 1993). Other patients have normal insulin action coupled with impaired insulin secretion (Alberti & Zimmet, 1998). Weight reduction, increased physical activity and the pharmacological treatment of hyperglycaemia have been reported to increase insulin sensitivity but not to its normal state (Wing *et al*., 1994). Unlike in T1D, ketoacidosis is infrequent and usually occurs as a result of stress or infection among people with T2D (Alberti & Zimmet, 1998). To avoid misclassifying T1D adults as T2D patients, detection of auto-antibodies is recommended (Alberti & Zimmet, 1998).

### 2.2.4 Other forms of diabetes

Other forms of diabetes are types of diabetes that have specific underlying causes, defects or disease processes that are used to identify them (Alberti & Zimmet, 1998). A number of underlying causes of the other forms of diabetes exist, encompassing specific genetic, infective or pharmacological causes, as indicated in Table 2.2 below.

**Table 2.2 Aetiological types of diabetes**

| |
|---|
| **Type 1** (beta-cell destruction, usually leading to absolute insulin deficiency)<br>Autoimmune, idiopathic |
| **Type 2** (may range from predominantly insulin-resistant with relative insulin deficiency to a predominantly secretory defect with or without insulin resistance) |
| **Other specific types**<br>MODY(maturity onset diabetes of the young)<br>Genetic defects of beta-cell function<br>Genetic defects in insulin action<br>Diseases of the exocrine pancreas<br>Endocrinopathies<br>Drug- or chemical-induced<br>Infections<br>Uncommon forms of immune-mediated diabetes<br>Other genetic syndromes sometimes associated with diabetes |
| Gestational diabetes |

Adapted from Alberti & Zimmet. (1998)

The other forms of diabetes have been reported to be rare and less prevalent compared to T1D and T2D (Alberti & Zimmet, 1998). T1D has also been noted to be less prevalent, accounting for less than 5% of diabetic patients (Amod *et al*., 2012). However, T2D is the most prevalent form of diabetes and it has been noted to affect more than 90% of the entire diabetic population. Therefore T2D is of more public health significance compared to other forms of diabetes and the rest of this chapter will focus on discussing the aspects of T2D (Amod *et al*., 2012).

## 2.3 METHODS FOR T2D DIAGNOSIS

The methods for diagnosing diabetes can be divided into two groups, consisting of either the determination of glucose concentration or glycated protein-based tests (Cox & Edelman, 2009). The glucose-based tests include fasting plasma glucose (FPG) and the OGTT (Cox & Edelman, 2009). Glycated haemoglobin (HbA1C) is an example of a glycated protein-based test that is used to diagnose T2D (IEC, 2009). The cut-off points for these tests are specified in Table 2.3 below.

**Table 2.3 Criteria for the diagnosis of diabetes**

| |
|---|
| 1. HbA1C > 6.5%. The test should be performed in a laboratory using a method that is NGSP-certified and standardised to the DCCT assay.* |
| OR |
| 2. FPG 126 $mgdL^{-1}$ (7.0 $mmolL^{-1}$). Fasting is defined as no caloric intake for at least 8 h.* |
| OR |
| 3. 2-h plasma glucose > 200 $mgdL^{-1}$ (11.1 $mmolL^{-1}$) during an OGTT. The test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water.* |
| OR |
| 4. In a patient with classic symptoms of hyperglycaemia or hyperglycaemic crisis, random plasma glucose 200 $mg.dL^{-1}$ (11.1 $mmolL^{-1}$). |

*In the absence of unequivocal hyperglycaemia, criteria 1–3 should be confirmed by repeat testing.*National Glycoheamoglobin Standardization Program (NGSP), Diabetes Control and Complications Trial (DCCT) assay. Adapted from ADA. (2013)

### 2.3.1 Fasting plasma glucose determination test

In the FPG test, glucose measurements are taken after a minimum of eight hours of fasting, which usually occurs overnight (Cox & Edelman, 2009). The FPG test has the advantages of being relatively easy and inexpensive to perform compared to other methods of diagnosing T2D. Lower intra-individual variability coefficients of 6.6-11.4% of FPG have been reported compared to 14.3%-16.7% for OGTT (Barr *et al.,* 2002). However, fasting can be an inconvenience and thus a negative aspect of the FPG method when compared to other methods such as the glycated haemoglobin test, which does not require people to fast. Processing of the blood sample has to be done within two hours to prevent erroneously low results and a second test is required to confirm the diagnosis (Cox & Edelman, 2009). In 2003, the FPG cut-off for diagnosing diabetes was reduced from 140 $mgdL^{-1}$ (7.8 $mmolL^{-1}$) to 126 $mgdL^{-1}$ (7 $mmolL^{-1}$) to improve sensitivity to diabetes that manifests as postprandial hyperglycaemia (Cox & Edelman, 2009). FPG correlates highly with diabetic retinopathy but is weakly associated with increased risk of mortality (Sorkin *et al.,* 2005).

### 2.3.2 Oral glucose tolerance test determination

The OGTT was introduced in 1922 and it is currently considered the gold standard for diagnosing diabetes (Cox & Edelman, 2009). IGT can only be diagnosed with the OGTT (Sacks *et al.,* 2002). The OGTT is more sensitive than the FPG test, as it diagnoses 2% more patients compared to the FPG test (Sacks *et al.,* 2002). However, the OGTT has poor reproducibility when compared to the FPG test and HbA1C measurements (Cox & Edelman, 2009). The eight-hour fast, commitment of nursing staff, length of the test and the necessity of an additional visit required for OGTT make this less convenient to perform (Cox & Edelman, 2009). The OGTT has good sensitivity and specificity of 87.5% and 75.8% respectively for identifying diabetic retinopathy (McCance *et al.,* 1994).

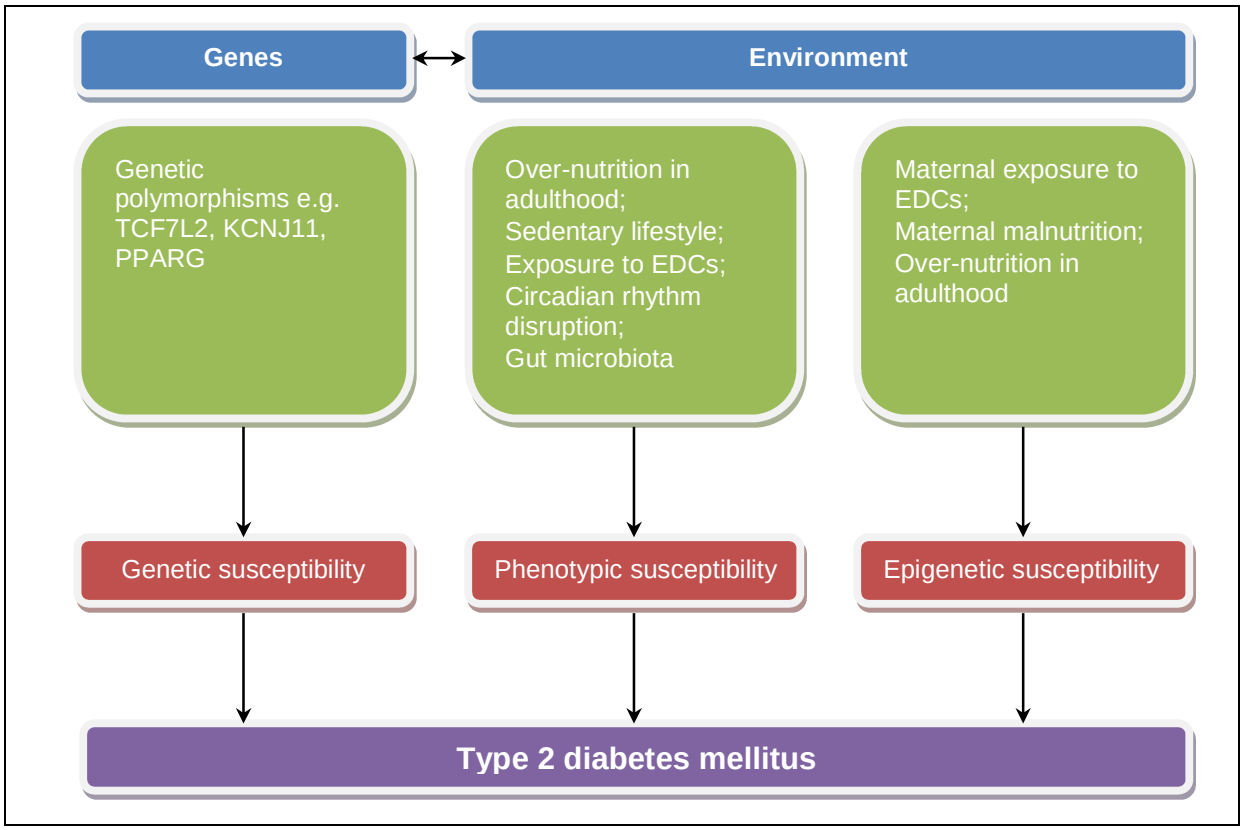
### 2.3.3 Glycated haemoglobin determination

In 1976, the HbA1C test was introduced as a measure for blood glucose control (Koenig *et al.,* 1976). Since then the HbA1C test has been accepted as the gold standard for glycaemic control and prognostication (Cox & Edelman, 2009). However, in 2009 the ADA endorsed the use of HbA1C for screening and diagnosing diabetes (Cox & Edelman, 2009). A cut-off level of HbA1C ≥ 6.5% was also issued by the International Expert Committee for diabetes diagnosis around the same time (IEC, 2009). A number of studies have been conducted to assess the specificity and sensitivity of HbA1C compared to FPG and OGTT (Cox & Edelman, 2009; Davidson *et al*., 1999). HbA1C was noted to have 83.4% sensitivity and 84.4% specificity of identifying diabetes when compared to FPG cut-off points (Davidson *et al*., 1999). Comparisons of OGTT and HbA1C testing indicated that HbA1C had 81.8% sensitivity and a specificity of 85% for diagnosing diabetes (Cox & Edelman, 2009). However, HbA1C has a low sensitivity of 30% for diagnosing IGT and is therefore not a better option than OGTT for IGT diagnosis (Cox & Edelman, 2009). HbA1C levels are strongly correlated with diabetic retinopathy and increased risk of cardiovascular diseases (Cox & Edelman, 2009). However, controversy persists regarding the suitability of the HbA1C test in South Africa, particularly in view of the high prevalence of iron deficiency anaemia, which affects the accuracy of this test (Amod *et al*., 2012).

## 2.4 FRAMEWORK FOR THE AETIOLOGY OF TYPE 2 DIABETES

T2D is a complex, multifactorial disease whose aetiology has not yet been fully elucidated (Goh, 2012). However, the complex interaction of the genetic and environmental factors from the gestational stages to adulthood, involving also epigenetic mechanisms, is hypothesised to result in the development of T2D (Goh, 2012). The gene-environment interactions have been hypothesised to span generations (Bollati & Baccarelli, 2010). T2D has a polygenic aetiology and unlike monogenic disorders where genetic susceptibility is due to a single genetic alteration with a high penetrance, in the case of T2D a number of genes with modest effects are postulated to interact with environmental factors to cause the disorder (Goh, 2012). Researchers now understand that the aetiology of complex diseases such as T2D may be better explained by a gene/epigene-environment framework, as illustrated in Figure 2.1 below (Goh, 2012). Epigenetics consists of the inherited gene expression patterns that occur without changes in the deoxyribonucleic acid (DNA) sequence (Rakyan & Whitelaw, 2003).

**Figure 2.1 Interaction of genetic and environmental factors in the development of T2D.**




EDCs=endocrine disrupting chemicals; Adapted from Goh. (2012)

### 2.4.1 Gene-environmental interactions in infancy

The maternal environment and feeding practices during infancy are key stages for T2D susceptibility (Goh, 2012). Intra-uterine growth retardation has been linked to diseases later in life such as T2D, obesity and hypertension (Barker, 2007). Evidence exists that the intrauterine environment of the diabetic mother confers risk of developing T2D and obesity in offspring (Goh, 2012). The multi-ethnic Search for Diabetes in Youth Study and longitudinal studies among the Pima Indians indicated that T2D during pregnancy was associated with early onset of the disease among the offspring later in life (Dabelea *et al.,* 2000; Franks *et al*., 2006; Pettit *et al.,* 2008). Certain micronutrient deficiencies, which include folate and vitamin B12, during pregnancy, were also linked to childhood adiposity and insulin resistance among Indians (Yajnik *et al*., 2008). Breastfeeding confers

protection from the development of T2D up to the age of 21 (Mayer-Davis *et al.,* 2008). Foetal and neonatal programming, which is hypothesised to occur as a result of epigenetic mechanisms, is postulated to be the mechanism by which the environmental factors early in life may lead to T2D susceptibility (Barker*,* 2007; Goh, 2012). However, most of the evidence for foetal programming has been reported in animal studies and more research in humans is required (Goh, 2012).

### 2.4.2 Diet, physical activity and T2D susceptibility

Later in life, from infancy to adulthood, there is constant exposure to a number of environmental factors that increase the risk of T2D (Goh, 2012). The global increase in T2D prevalence is hypothesised to be attributed to the adoption of Westernised lifestyles among people who are genetically and epigenetically susceptible to T2D (Goh, 2012). T2D was once viewed as a Western disease and a disease of the affluent; however, studies reveal that the disease is now a global burden affecting the poor as well (WHO, 2008). The Western lifestyle is characterised by the consumption of energy-dense foods and reduced physical activity (Franks, 2013). Globalisation and rapid urbanisation, which are occurring worldwide, are postulated to be catalysing the adoption of the Western lifestyle (Franks, 2013). The rates of obesity and T2D have been noted to rise sharply as communities move from rural to urban environments (O'Dea *et al.,* 1991). Marketing and the affordability of foods high in sugar and fat in developing countries is aiding the nutrition transition from traditional foods protective against T2D to those that increase its risk (Goh, 2012). Micronutrient imbalances of vitamin D and B12 in patients replete in folic acid and who have increased iron stores have been linked to increased risk of developing T2D (Barret *et al.,* 2010). Technology enhancements have reduced physical activity and thus energy expenditure, making people more prone to obesity, which is a strong risk factor in the development of T2D (Dunstan *et al.,* 2007). Urbanisation has also made sedentary lifestyles accessible to many and therefore it has been associated with obesity and T2D (Dunstan *et al.,* 2007). The sudden increased migration to urban environments characterised by high energy intake by people who are epigenetically programmed and genetically susceptible to favour a low energy intake is postulated to be the mechanism

underlying the increased T2D risk, as a result of the adoption of Western lifestyles (Hochberg *et al.,* 2010).

### 2.4.3 Microbiome and T2D susceptibility

The gut microbiota has also been linked to increased T2D risk (Goh, 2012). The microbiome of the obese has been determined to increase the gut's ability to capture energy from the diet (Turnbaugh *et al*., 2006). The increased fermentation capacity of the microbiota generates more short chain fatty acids, which are linked to the development of obesity through lipogenesis (Greiner *&* Bäckhed*,* 2011). The change in the proportions of the gut microbiota among T2D patients was reported to be correlated with glucose levels (Larsen *et al*., 2010). However, more research is required to elucidate the precise mechanism by which changes in the gut microbiota increase T2D susceptibility (Goh, 2012).

### 2.4.4 Endocrine-disrupting chemicals and T2D susceptibility

Endocrine-disrupting chemicals (EDCs) are defined as exogenous agents that hinder the synthesis, secretion, transport and metabolism of hormones responsible for homeostasis, reproduction and developmental processes (Diamanti-Kandarakis *et al*., 2009). The effects of EDCs on increasing disease susceptibility span generations in animal models (Skinner *et al*., 2010). It is postulated that EDCs interact with the genome and/or the epigenome to exert transgenerational effects of increased disease susceptibility (Skinner *et al*., 2010). Exposure to organic pollutants, which include pesticides and plasticisers, has been reported to disrupt the functioning of the endocrine cells, leading to increased risk of developing T2D (Diamanti-Kandarakis *et al*., 2009). Bisphenol (BPA), an oestrogenic compound used as a plasticiser, was reported to cause temporary hyperinsulinaemia and in the long term, insulin resistance (Alonso-Magdalena *et al*., 2006). Adiponectin, a hormone that enhances insulin sensitivity, was reported to be inhibited by even low doses of BPA (Hugo *et al*., 2008). Some countries, which include Canada, banned the use of BPA in baby bottles because of its potentially adverse effects (Goh, 2012). Dioxin and related compounds, which are present in petroleum-derived herbicides, have been noted to

reduce the production of the glucose transporter type 4, resulting in progressive insulin resistance (Remillard & Bunce, 2002). Advanced glycation end products (AGEs) are produced during cooking at high temperatures and increase oxidative stress, a key component of T2D pathogenesis through the production of reactive oxygen species (Sandu *et al.,* 2005). High-AGE diets are associated with the development of T2D and obesity in animal models (Sandu *et al.,* 2005).

In summary, T2D is a complex disease, which requires pre-existing genetic susceptibility for its development and significant environmental exposure to trigger its clinical expression (Goh, 2012).

## 2.5 HERITABILITY OF TYPE 2 DIABETES

T2D has been noted to have a strong genetic component to its aetiology (Florez, 2013). Evidence of the genetic component of T2D aetiology has been derived from twin studies, family-based studies and heritability estimates (Florez, 2013). A high concordance rate ranging up to 96% has been reported among monozygotic twins, whereas that of dizygotic twins was about 20-30% (Newman *et al.,* 1987; Medici *et al*., 1999). Through family-based genetic studies, it was noted that a 40% risk of developing T2D exists for children of diabetic parents, with the risk increasing to 70% if both parents have the disorder (Köbberling & Tillil, 1982). The sibling relative risk ($\lambda_S$), a measure used to compare the likelihood of developing a disease between the relatives of a patient and the general population, was noted to be 3 for T2D in various population groups, indicative of a strong heritable component (Lyssenko *et al*., 2005). On average the heritability ($h^2$) estimates of T2D are about 0.49, implying that genetic factors make a contribution of almost 50% to the development of this disease (Risch, 1990). However, the twin studies and heritability estimates do not clarify the precise factors that make up the genetic aetiology of T2D (Goh, 2012). Various methods discussed below have thus been adopted to decipher the genetic aetiology of T2D.

### 2.5.1 Linkage analysis

Linkage analysis studies among the affected families were one of the earliest methods used to study the genetic variants associated with disease susceptibility (Groop & Pociot, 2013). Long stretches of linkage disequilibrium (LD) were analysed by genotyping 400-500 genetic markers throughout the genome to map disease loci on a genome-wide level (Ahlqvist *et al*., 2010). The shared markers among the affected family members compared to the unaffected individuals indicated that a disease-causing variant was most likely in LD with a specific genotyped marker (Ahlqvist *et al*., 2010). Linkage analysis enabled the discovery of many of the genetic susceptibility loci of simple Mendelian (monogenic) diseases (Sanghera & Blackett, 2012). However, it was noted to be less useful for deciphering the genetic aetiology of complex diseases such as T2D (Groop & Pociot, 2013). Only two genes, calpain 10 (CAPN10) and transcription factor 7-like 2 (TCF7L2), were discovered through linkage analysis studies for T2D. TCF7L2 was later associated with T2D in various ethnicities, including black South Africans of Zulu descent, and to date its risk allele at the rs7903146 SNP has the strongest association with T2D (Tong *et al.,* 2009; Pirie *et al.,* 2010). However, none of the later genome-wide association methods was able to replicate the association of CAPN10 with T2D (Groop & Pociot, 2013). After the linkage analysis, the candidate gene association approach was considered to unravel the genetic aetiology of T2D.

### 2.5.2 Candidate gene association studies for type 2 diabetes

Unlike linkage analysis studies, which are family-based, candidate gene studies are population-based but also rely on the mapping of LD stretches using genetic markers to map disease susceptibility loci (Groop *&* Pociot., 2013). However, more markers are used in candidate gene association studies because of the short LD stretches of unrelated individuals, to map the genetic loci at the genome-wide level (Ahlqvist *et al.,* 2011). Candidate genes are selected based on function and their greater probability of affecting the disease trait (Ahlqvist *et al.,* 2011). Peroxisome proliferator-activated receptor gamma (*PPARG*), insulin receptor substrate 1 (*IRS1*), potassium inwardly-rectifying channel, subfamily J, member 11 (*KCNJ11*), Wolfram syndrome 1 (*WFS1*), HNF1 homeobox A

(*HNF1A*) and HNF1 homeobox B (*HNF1B*) are the six candidate genes that have been consistently associated with T2D (Ahlqvist *et al.,* 2011). Most of the other candidate genes were not replicated in other populations and inconsistent results have been reported (Zeggini *et al.,* 2007; Barroso *et al*., 2003; Elbein, 1997). It was noted that incorrect candidate gene selection due to inadequate knowledge about the genetic aetiology of the disease, heterogeneity of the disease, irregularities in study designs, over-interpretation of results and population-specific LD were some of the reasons for the failure of linkage analysis and candidate gene association studies to decipher the genetic aetiology of complex diseases such as T2D (Zeggini *et al.,* 2007; Barroso *et al*., 2003; Elbein, 1997). The GWAS approach was the next method after the candidate gene association approach to be used to decipher T2D genetic susceptibility.

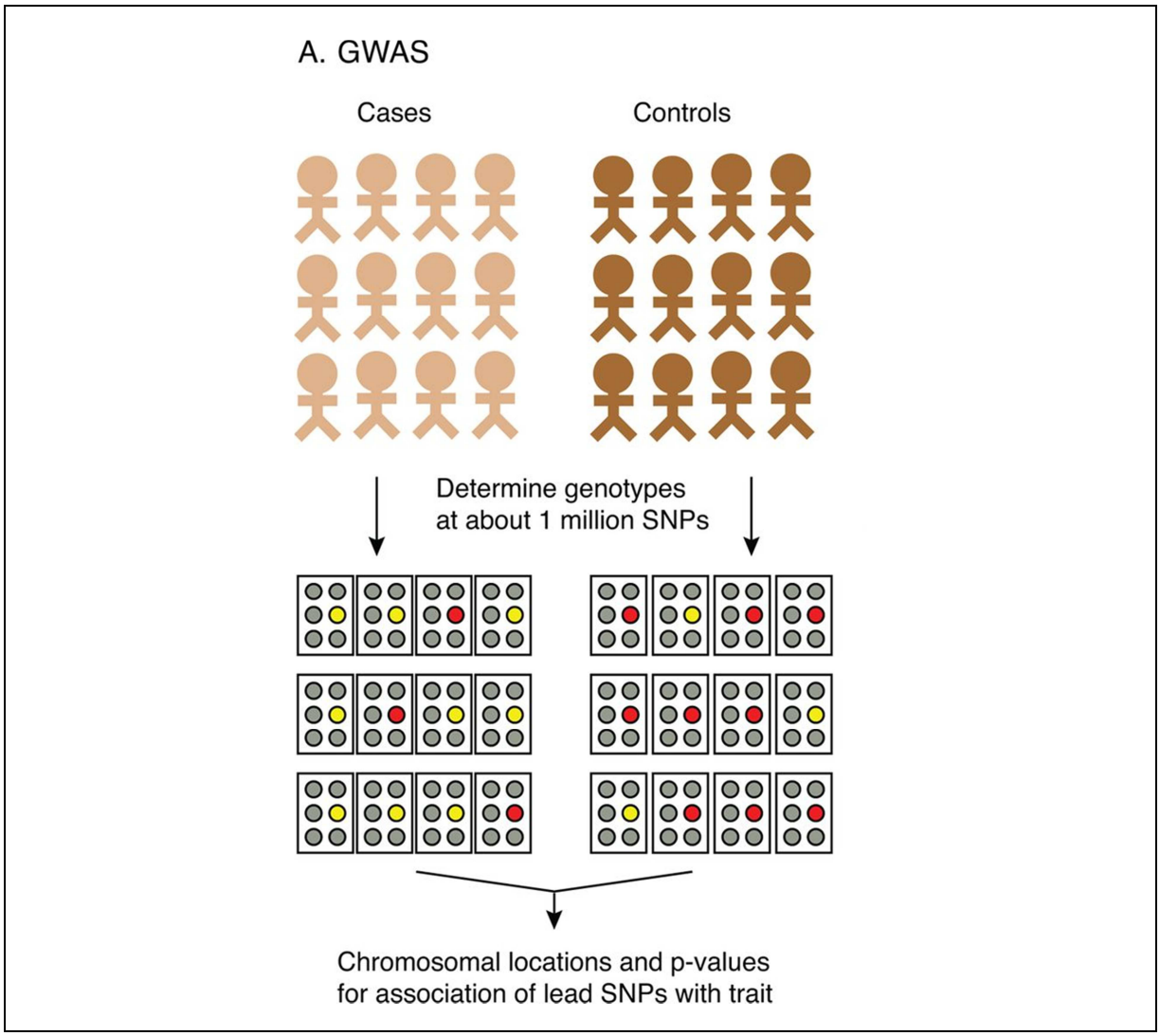
### 2.5.3 Genome-wide association studies and type 2 diabetes genetic susceptibility

GWASs, similar to earlier methods such as linkage analysis and candidate gene association studies, rely on LD to map disease susceptibility loci (Bush and Moore, 2012). However, the genetic markers in the GWAS are termed SNPs, which are single base-pair changes in the DNA sequence (Genomes Project Consortia, 2010). SNPs are postulated to be the most common form of genetic variation in the human genome (Bush & Moore, 2012). The common disease/common variant hypothesis (CD/CV) forms the basis for use of SNPs as markers for disease susceptibility loci in GWAS (Bush & Moore, 2012). The CD/CV hypothesis states that common diseases are influenced by genetic variation, which is common throughout the human population (Reich & Lander, 2001). Thus, the genetic susceptibility of complex and common disease such as T2D was among the first to be evaluated using the GWAS approach (Sanghera & Blackett, 2012).

Major changes in the genetic research landscape made GWAS feasible (Florez, 2013). The complete sequencing of the human genome, the cataloguing of the human genetic variation in the HapMap project and rapid improvements in high throughput sequencing and genotyping technology made the simultaneous genotyping of hundreds of thousands of SNPs feasible (Groop & Pociot, 2013). The HapMap project led to the understanding

that 500 000 to 1 000 000 tagging SNPs could be used to capture more than 80% of the common variation in the human genome (Li *et al*., 2008). GWAS designs can have either case control or quantitative study designs (Bush & Moore, 2012). The quantitative study designs for GWAS are deemed suitable when the disease in question has an established biomarker that is measurable, such as low density lipoprotein in the case of cardiovascular diseases (Bush & Moore, 2012). In quantitative designs, the genetic variants that affect the levels of the biomarkers are evaluated (Bush & Moore, 2012). However, in case control study designs the established disease phenotypes are identified and the genetic variants associated with the phenotype are assessed by comparing the distribution patterns of the selected genetic variants as depicted in Figure 2.2 (Bush & Moore, 2012). The SNPs determined to be associated with phenotypic traits via a GWAS are deemed not to be causal (Bush & Moore, 2012). The SNPs used in GWAS are usually tagging SNPs, which are hypothesised to be in LD with the causal SNP (Hirschhorn & Daly, 2005). Additional studies for fine mapping are performed later to identify the causal SNPs (Bush & Moore, 2012). Therefore GWAS findings provide indirect association of genetic variants with disease phenotypes or quantitative traits (Hirschhorn & Daly, 2005).

**Figure 2.2 GWAS case control study design**




Adapted from Hardison. (2012)

### 2.5.4 Genome-wide association studies of type 2 diabetes in European populations

The earliest GWAS for T2D revealed haematopoietically expressed homeobox (*HHEX*), and solute carrier family 30 (zinc transporter), member 8 (*SLC30A8*) as new diabetes loci and replicated the association of *TCF7L2* and *KCNJ11,* which had previously been discovered through linkage analysis and candidate gene association studies respectively (Sladek *et al.,* 2007). In a later analysis, three GWASs, which were based in European

individuals, allowed for the discovery of the *CDK5* regulatory subunit-associated protein 1-like 1 (*CDKAL1*) gene and a variant near cyclin-dependent kinase inhibitor 2A (*CDKN2A-B*) as novel T2D loci (Zeggini *et al*., 2008; Scott *et al*., 2007). Confirmation of the association of *TCF7L2, KCNJ11, PPARG, SLC30A8* and *HHEX* loci with T2D was also noted in the three European GWASs (Zeggini *et al*., 2008; Scott *et al*., 2007). The fat mass and obesity-associated (*FTO*) gene was also determined to be associated with T2D risk through its effect on obesity (Scott *et al.,* 2007). Different GWASs were combined into data sets in meta-analysis studies with the view to increase the power of detecting other disease susceptibility loci (Qi & Hu, 2012). The Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium was developed and it was able to discover six additional loci associated with T2D risk, which comprised JAZF zinc finger 1 (*JAZF1*), cell division cycle 123 (*CDC123*)/calcium/calmodulin-dependent protein kinase ID (*CAMK1D*), tetraspanin 8 (*TSPAN8)/* leucine-rich repeat containing G protein-coupled receptor 5 (*LGR5*), thyroid adenoma associated (*THADA),* ADAM metallopeptidase with thrombospondin type 1 motif, 9 (*ADAMTS9)*, and notch 2 *(NOTCH2)* in 10 128 individuals, using 2.2 million SNPs (Zeggini *et al.,* 2008). Later in 2009, *IRS1* and melatonin receptor 1B (*MTNR1B*) were identified through GWAS as loci for T2D risk (Bouatia-Naji *et al*., 2009; Rung *et al*., 2009). In 2010, the DIAGRAM consortium was complemented with other GWAS cohorts from the deCODE genetics group, the Diabetes Gene Discovery group, the KORA group, the Rotterdam study and the EUROSPAN to form the DIAGRAM plus consortium, a meta-analysis study which had a sample size of 22 044 participants (Voight *et al.,* 2010). Twelve novel loci associated with T2D risk were discovered from the DIAGRAM plus meta-analysis, which included B-cell CLL/lymphoma 11A, zinc finger protein *(BCL11A),* zinc finger, BED-type containing 3 *(ZBED3),* Kruppel-like factor 14 *(KLF14),* tumour protein p53 inducible nuclear protein 1 (*TP531NP1)*, transducin-like enhancer of split 4 (E(sp1) homologue, Drosophila) *(TLE4),* potassium voltage-gated channel, KQT-like subfamily, member 1 (*KCNQ1*), high mobility group AT-hook 2 *(HMGA2), HNF1A,* zinc finger, AN1-type domain 6 *(ZFAND6),* protein regulator of cytokinesis 1 *(PRC1),* dual specificity phosphatase 9 *(DUSP9)* and ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1 *(ARAP1)* (Voight *et al.,* 2010). In 2010, 21 GWASs were combined to form the Meta-

analysis of Glucose and Insulin-related Traits Consortium (MAGIC) group (Qi & Hu, 2012). In the MAGIC study, 18 loci associated with fasting glucose and/or insulin were discovered (Dupuis *et al*., 2010). However, only five of the loci identified in the MAGIC study were noted to be associated with T2D risk; these were adenylate cyclase 5 *(ADCY5),* prospero homeobox 1 *(PROX1),* glucokinase, hexokinase 4 *(GCK),* glucokinase (hexokinase 4) regulator *(GCKR)* and diacylglycerol kinase, beta 90kDa-alkylglycerol monooxygenase *(DGKB-TMEM195)* (Dupuis *et al*., 2010). To date 39 loci have been discovered to be associated with T2D through GWASs in European populations (Qi & Hu, 2012).

### 2.5.5 Genome-wide association studies of type 2 diabetes in other populations

GWASs of T2D in other ethnicities were not able to replicate all the variants that had been identified in the European populations (Sanghera & Blackett, 2012). However, other novel SNPs were also discovered in these ethnicities (Sanghera & Blackett, 2012). GWASs among East Asians identified a new locus, namely ubiquitin-conjugating enzyme E2E 2 (*UBE2E2),* associated with T2D, and follow-up meta-analyses were able to identify GLIS family zinc finger 3 (*GLIS3*), peptidase D (*PEPD*), fat storage-inducing transmembrane protein 2-R3H domain containing-like-hepatocyte nuclear factor 4, alpha (*FITM2-R3HDML-HNF4A*), potassium channel, subfamily K, member 16 (*KCNK16*), macrophage erythroblast attacher (*MAEA*), GRIP and coiled-coil domain containing 1- paired box 4 (*GCC1-PAX4*), proteasome (prosome, macropain) 26S subunit, non-ATPase, 6 (*PSMD6*) and zinc finger, AN1-type domain 3 (*ZFAND3*) as eight more novel loci (Yasuda *et al*., 2008; Cho *et al.,* 2011). Among the South Asians, six novel loci were also discovered to be associated with T2D, which included growth factor receptor-bound protein 14 (*GRB14*), ST6 beta-galactosamide alpha-2, 6-sialyltranferase 1 (*ST6GAL1*), vacuolar protein sorting 26 homolog A, *S. pombe* (*VPS26A*), high mobility group 20A (*HMG20A*), adaptor-related protein complex 3, sigma 2 subunit (*AP3S2*) and hepatocyte nuclear factor 4, alpha (*HNF4A*) (Kooner *et al*., 2008). Although 96% of the GWASs for T2D were performed in European populations, identification of novel variants in genes such as *KCNQ1* among Japanese, which later on were associated with T2D in Europeans, indicated that one population group is inadequate to provide information on T2D genetic susceptibility for the

global population (Sanghera & Blackett, 2012). The *KCNQ1* locus, which was first discovered among the Japanese, also provided new insights into the aetiology of T2D, particularly the pathways for insulin resistance (Yasuda *et al*., 2008). The genetic diversity that has been noted to exist in people of different ethnicities has led to the hypothesis that even the best designed GWAS in one population is not sufficiently representative to allow for the evaluation of genetic risk in another population (Sanghera & Blackett, 2012). To date, about 79 common variants have been associated with T2D through the GWAS approach in different ethnicities and these are depicted in Table 2.4 (Sanghera & Blackett, 2012).

**Table 2.4 The list of SNPS that have been associated with T2D in various ethnicities through GWAS**

| *SNP* | *Gene* | *Gene Region* | *Population* | *Trait* | *MAF in Controls* | *OR[95%]* | *P-Value* | *Reference* |
|---|---|---|---|---|---|---|---|---|
| **rs10923931** | *NOTCH2,ADAM30* | 1p12 | European | T2D | T [0.11] | 1.13 [1.08–1.17] | $4\times10^{-8}$ | Zeggini *et al.,* 2008 |
| **rs243021** | *BCL11A* | 2p16.1 | European | T2D | A [NR] | 1.08 [1.06–1.10] | $3\times10^{-15}$ | Voight *et al*., 2010 |
| **rs7578326** | *IRS1* | 2q36.3 | European | T2D | A [NR] | 1.11 [1.08–1.13] | $5\times10^{-20}$ | Voight *et al.,* 2010 |
| **rs2943641** | *LOC64673,IRS1* | 2q36.3 | European | T2D/ Other | C [0.63] | 1.19 [1.13–1.25] | $9\times10^{-12}$ | Rung *et al.,* 2009 |
| **rs7593730** | *RBMS1,ITGB6* | 2q24.2 | European | T2D | T [0.78] | 1.11 [1.08–1.16] | $4\times10^{-8}$ | Qi *et al*., 2010 |
| **rs7578597** | *THADA* | 2p21 | European | T2D | T [0.90] | 1.15 [1.10–1.20] | $1\times10^{-9}$ | Zeggini *et al*., 2008 |
| **rs4607103** | *ADAMTS9* | 3p14.1 | European | T2D | C [0.76] | 1.09 [1.06–1.12] | $1\times10^{-8}$ | Zeggini *et al*., 2008 |
| **rs4402960** | *IGF2BP2* | 3q27.2 | European | T2D | T [0.30] | 1.14 [1.11–1.18] | $9\times10^{-16}$ | Scott *et al*., 2007 |
| **rs358806** | *NR* | 3p14 | European | T2D | A [0.80] | 1.16 [1.03–1.33] | $3\times10^{-6}$ | WTCC Consortium, 2007 |
| **rs1801282** | *PPARG* | 3p25.2 | European | T2D | C [0.82] | 1.14 [1.08–1.20] | $2\times10^{-6}$ | Scott *et al*., 2007 |
| **rs17036101** | *SYN2,PPARG* | 3p25.2 | European | T2D | G [0.93] | 1.15 [1.10–1.21] | $2\times10^{-7}$ | Zeggini *et al*., 2008 |
| **rs1801214** | *WFS1* | 4p16.1 | European | T2D | T [NR] | 1.13 [1.08–1.18] | $3\times10^{-8}$ | Voight *et al*., 2010 |
| **rs12518099** | *LOC72901,CETN3* | 5q14.3 | European | T2D/ Other | C [0.23] | 1.16 [1.10–1.22] | $7\times10^{-7}$ | Rung *et al*., 2009 |
| **rs4457053** | *ZBED3* | 5q13.3 | European | T2D | G [NR] | 1.08 [1.06–1.11] | $3\times10^{-12}$ | Voight *et al*., 2010 |
| **rs7754840** | *CDKAL1* | 6p22.3 | European | T2D | C [0.36] | 1.12 [1.08–1.16] | $4\times10^{-11}$ | Scott *et al*.,2007 |
| **rs896854** | *TP53INP1* | 6q22.1 | European | T2D | T [NR] | 1.06 [1.04–1.09] | $1\times10^{-9}$ | Voight *et al*., 2010 |
| **rs9472138** | *VEGFA* | 6p21.1 | European | T2D | T [0.28] | 1.06 [1.04–1.09] | $4\times10^{-6}$ | Zeggini *et al*., 2008 |
| **rs864745** | *JAZF1* | 7p15.1 | European | T2D | T [0.5] | 1.10 [1.07–1.13] | $5\times10^{-14}$ | Zeggini *et al*., 2008 |
| **rs972283** | *KLF14* | 7q32.3 | European | T2D | G [NR] | 1.07 [1.05–1.10] | $2\times10^{-10}$ | Voight *et al*., 2010 |
| **rs13266634** | *SLC30A8* | 8q24.11 | European | T2D | C [0.61] | 1.12 [1.07–1.16] | $5\times10^{-8}$ | Scott *et al*., 2007 |
| **rs564398** | *CDKN2A,CD-KN2B* | 9p21.3 | European | T2D | T [0.56] | 1.13 [1.08–1.19] | $1\times10^{-6}$ | Zeggini *et al*., 2007 |
| **rs7018475** | *CDKN2B* | 9p21.3 | European | T2D | G [NR] | 1.35 [1.18–1.56] | $3\times10^{-8}$ | Huang *et al*., 2012 |
| **rs13292136** | *CHCHD9* | 9q21.31 | European | T2D | C [NR] | 1.11 [1.07–1.15] | $3\times10^{-8}$ | Voight *et al*., 2010 |
| **rs12779790** | *CDC123,CAMK1D* | 10p13 | European | T2D | G [0.18] | 1.11 [1.07–1.14] | $1\times10^{-10}$ | Zeggini *et al*.,2008 |
| **rs1111875** | *HHEX* | 10q23.33 | European | T2D | C [0.52] | 1.13 [1.09–1.17] | $6\times10^{-10}$ | Scott *et al*., 2007 |
| **rs1552224** | *ARAP1* | 11q13.4 | European | T2D | A [NR] | 1.14 [1.11–1.17] | $1\times10^{-22}$ | Voight *et al*., 2007 |
| **rs5219** | *KCNJ11* | 11p15.1 | European | T2D | T [0.46] | 1.14 [1.10–1.19] | $7\times10^{-11}$ | Scott *et al.,* 2007 |
| **rs231362** | *KCNQ1* | 11p15.5 | European | T2D | G [NR] | 1.08 [1.06–1.10] | $3\times10^{-13}$ | Voight *et al.,* 2010 |
| **rs1387153** | *MTNR1B* | 11q14.3 | European | T2D | T [NR] | 1.09 [1.06–1.11] | $8\times10^{-15}$ | Voight *et al.,* 2010 |

| SNP | Gene | Gene Region | Population | Trait | MAF in Controls | OR[95%] | P-Value | Reference |
|---|---|---|---|---|---|---|---|---|
| **rs1153188** | *DCD* | 12q13.2 | European | T2D | A [0.73] | 1.08 [1.05–1.11] | $2\times10^{-7}$ | Zeggini *et al.,* 2008 |
| **rs1531343** | *HMGA2* | 12q14.3 | European | T2D | C [NR] | 1.10 [1.07–1.14] | $4\times10^{-9}$ | Voight *et al.,* 2010 |
| **rs7957197** | *HNF1A* | 12q24.31 | European | T2D | T [NR] | 1.07 [1.05–1.10] | $2\times10^{-8}$ | Voight *et al.,* 2010 |
| **rs7961581** | *TSPAN8,LGR5* | 12q21.1 | European | T2D | C [0.27] | 1.09 [1.06–1.12] | $1\times10^{-9}$ | Zeggini *et al.,* 2008 |
| **rs8042680** | *PRC1* | 15q26.1 | European | T2D | A [NR] | 1.07 [1.05–1.09] | $2\times10^{-10}$ | Voight *et al.,* 2010 |
| **rs11634397** | *ZFAND6* | 15q25.1 | European | T2D | G [NR] | 1.06 [1.04–1.08] | $2\times10^{-9}$ | Voight *et al.,* 2010 |
| **rs8050136** | *FTO* | 16q12.2 | European | T2D | A [0.40] | 1.23 [1.18–1.32] | $7\times10^{-14}$ | Zeggini *et al*., 2007 |
| **rs4430796** | *HNF1B* | 17q12 | European | T2D | G [NR] | 1.14 [1.08–1.20] | $2\times10^{-6}$ | Voight *et al.,* 2010 |
| **rs5945326** | *DUSP9* | Xq26 | European | T2D | G [NR] | 1.27 [1.18–1.37] | $3\times10^{-10}$ | Voight *et al.,* 2010 |
| **rs7542900** | *SLC44A3,F3* | 1p21.3 | Afr. Amer. | T2D | C [0.56] | 1.16 [1.09–1.25] | $6\times10^{-6}$ | Palmer *et al*., 2012 |
| **rs7560163** | *RBM43,RND3* | 2q23.3 | Afr. Amer. | T2D | C [0.86] | 1.33 [1.19–1.49] | $7\times10^{-9}$ | Palmer *et al., 2012* |
| **rs2722769** | *GALNTL4, LOC729013* | 11p15.3 | Afr. Amer. | T2D | C [0.53] | 1.35 [1.19–1.54] | $2\times10^{-6}$ | Palmer *et al., 2012* |
| **rs7107217** | *TMEM45B, BARX2* | 11q24.3 | Afr Amer. | T2D | C [0.91] | 1.18 [1.10–1.27] | $3\times10^{-7}$ | Palmer *et al.,* 2012 |
| **rs17045328** | *CR2* | 1q32.2 | Asian | T2D | G [0.30] | 1.38 [1.20–1.59] | $7\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs12027542** | *PCNXL2* | 1q42.2 | Asian | T2D | A [0.61] | 1.41 [1.23–1.61] | $4\times10^{-7}$ | Sim *et al.,* 2011 |
| **rs7630877** | *PEX5L* | 3q26.33 | Asian | T2D | A [0.17] | 1.32 [1.17–1.49] | $7\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs3773506** | *PLS1* | 3q23 | Asian | T2D | C [0.06] | 1.81 [1.39–2.35] | $9\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs831571** | *PSMD6* | 3p14.1 | Asian | T2D | C [0.61] | 1.09 [1.06–1.12] | $8\times10^{-11}$ | Cho *et al*., 2011 |
| **rs2063640** | *ZPLD1* | 3q12.3 | Asian | T2D | A [0.17] | 1.23 [1.13–1.34] | $3\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs6815464** | *MAEA* | 4p16.3 | Asian | T2D | C [0.58] | 1.13 [1.10–1.16] | $2\times10^{-20}$ | Cho *et al*., 2011 |
| **rs1048886** | *C6orf57* | 6q13 | Asian | T2D | G [0.18] | 1.54 [1.32–1.80] | $3\times10^{-8}$ | Sim *et al.,* 2011 |
| **rs1535500** | *KCNK16* | 6p21.2 | Asian | T2D | T [0.42] | 1.08 [1.05–1.11] | $2\times10^{-8}$ | Cho *et al.,* 2011 |
| **rs9470794** | *ZFAND3* | 6p21.2 | Asian | T2D | C [0.27] | 1.12 [1.08–1.16] | $2\times10^{-10}$ | Cho *et al*., 2011 |
| **rs7636** | *ACHE* | 7q22.1 | Asian | T2D | A [0.06] | 1.85 [1.42–2.41] | $5\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs7041847** | *GLIS3* | 9p24.2 | Asian | T2D | A [0.41] | 1.10 [1.07–1.13] | $2\times10^{-14}$ | Cho *et al*., 2011 |
| **rs17584499** | *PTPRD* | 9p24.1 | Asian | T2D | T [0.06] | 1.57 [1.36–1.82] | $9\times10^{-10}$ | Tsai *et al*., 2010 |
| **rs6583826** | *KIF11* | 10q23.33 | Asian | T2D | G [0.26] | 1.18 [1.10–1.27] | $7\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs10741243** | *TCERG1L* | 10q26.3 | Asian | T2D | G [0.93] | 1.75 [1.38–2.23] | $5\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs1359790** | *SPRY2* | 13q31.1 | Asian | T2D | G [0.71] | 1.15 [1.10–1.20] | $6\times10^{-9}$ | Shu *et al*., 2010 |
| **rs1436955** | *C2CD4B* | 15q22.2 | Asian | T2D | C [0.73] | 1.13 [1.08–1.19] | $7\times10^{-7}$ | Shu *et al*., 2010 |

| SNP | Gene | Gene Region | Population | Trait | MAF in Controls | OR[95%] | P-Value | Reference |
|---|---|---|---|---|---|---|---|---|
| **rs16955379** | *CMIP* | 16q23.2 | Asian | T2D | C [0.80] | 1.08 [1.05–1.12] | $3\times10^{-7}$ | Cho *et al*., 2011 |
| **rs17797882** | *WWOX* | 16q23.2 | Asian | T2D | T [0.32] | 1.08 [1.05–1.12] | $9\times10^{-7}$ | Cho *et al*., 2011 |
| **rs391300** | *SRR* | 17p13.3 | Asian | T2D | G [0.62] | 1.28 [1.18–1.39] | $3\times10^{-9}$ | Tsai *et al*., 2010 |
| **rs10460009** | *LPIN2* | 18p11.31 | Asian | T2D | C [0.60] | 1.35 [1.18–1.54] | $9\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs472265** | *PAPL* | 19q13.2 | Asian | T2D | G [0.22] | 1.39 [1.20–1.61] | $9\times10^{-6}$ | Sim *et al*., 2011 |
| **rs2833610** | *HUNK* | 21q22.11 | Asian | T2D | A [0.57] | 1.17 [1.09–1.24] | $4\times10^{-6}$ | Sim *et al.,* 2011 |
| **rs3923113** | *GRB14* | 2q24.3 | South Asian | T2D | A [0.74] | 1.09 [1.06–1.13] | $1\times10^{-8}$ | Kooner *et al*., 2011 |
| **rs16861329** | *ST6GAL1* | 3q27.3 | South Asian | T2D | G [0.75] | 1.09 [1.06–1.12] | $3\times10^{-8}$ | Kooner *et al*., 2011 |
| **rs1802295** | *VPS26A* | 10q22.1 | South Asian | T2D | A [0.26] | 1.08 [1.05–1.12] | $4\times10^{-8}$ | Kooner *et al*., 2011 |
| **rs2028299** | *AP3S2* | 15q26.1 | South Asian | T2D | C [0.31] | 1.10 [1.07–1.13] | $2\times10^{-11}$ | Kooner *et al*., 2011 |
| **rs7178572** | *HMG20A* | 15q24.3 | South Asian | T2D | G [0.52] | 1.09 [1.06–1.12] | $7\times10^{-11}$ | Kooner *et al*., 2011 |
| **rs4812829** | *HNF4A* | 20q13.12 | South Asian | T2D | A [0.29] | 1.09 [1.06–1.12] | $3\times10^{-10}$ | Kooner *et al*., 2011 |
| **rs2237897** | *KCNQ1* | 11p15.4 | Japanese | T2D | C [0.61] | 1.33 [1.24–1.41] | $1\times10^{-16}$ | Unoki *et al*., 2008 |
| **rs2237892** | *KCNQ1* | 11p15.5 | Japanese | T2D | C [0.61] | 1.40 [1.34–1.47] | $2\times10^{-42}$ | Yasuda *et al*., 2008 |
| **rs7172432** | *C2CD4A,C2CD4B* | 15q22.2 | Japanese | T2D | G [0.58] | 1.11 [1.08–1.14] | $9\times10^{-14}$ | Yamauchi *et al*., 2010 |
| **rs730570** | *C14orf70* | 14q32.2 | Mexican | T2D | G [NR] | 1.14 [1.08–1.21] | $8\times10^{-6}$ | Parra *et al*., 2010 |

Adapted from Sanghera & Blackett. (2012); NR-not reported; p = short arm of the chromosome; q = long arm of the chromosome; T2D = type 2 diabetes; Afr. Amer. = African American; MAF =minor allele frequency; *NOTCH2* = notch 2; *ADAM30* =ADAM metallopeptidase domain 30; BCL11A = B-cell CLL/lymphoma 11A (zinc finger protein); IRS = insulin receptor substrate 1; *RBMS1* = RNA binding motif; single stranded interacting protein 1; *ITGB6* = integrin; beta 6; *THADA* = thyroid adenoma associated; *ADAMTS9* = ADAM metallopeptidase with thrombospondin type 1 motif; 9; *IGF2BP2* = insulin-like growth factor 2 mRNA binding protein 2; *PPARG* = peroxisome proliferator-activated receptor gamma; *SYN2* = synapsin II; *WFS1* = Wolfram syndrome 1 (wolframin); *CETN3* = centrin; EF-hand protein; 3; *ZBED3* = zinc finger; BED-type containing 3; *CDKAL1* = CDK5 regulatory subunit associated protein 1-like 1; *TP53INP1* = tumor protein p53 inducible nuclear protein; *VEGFA* = vascular endothelial growth factor A; *JAZF1* = JAZF zinc finger 1; *KLF14* = Kruppel-like factor 14; *SLC30A8* = solute carrier family 30 (zinc transporter) member 8; *CDKN2A* = cyclin-dependent kinase inhibitor 2A; *CDKN2B* = cyclin-dependent kinase inhibitor 2B (p15; inhibits CDK4; *CHCHD9* = coiled-coil-helix-coiled-coil-helix domain containing 2 pseudogene 9; CDC123 = cell division cycle 123; CAMK1D = calcium/calmodulin-dependent protein kinase ID; HHEX = hematopoietically expressed homeobox; *ARAP1* = ArfGAP with RhoGAP domain; ankyrin repeat and PH domain 1; *KCNJ11*= potassium inwardly-rectifying channel; subfamily J; member 11; *KCNQ1* = potassium voltage-gated channel; KQT-like subfamily; member 1; *MTNR1B* = melatonin receptor 1B; *DCD* = dermcidin; *HMGA2* = high mobility group AT-hook 2; *HNF1A* = HNF1 homeobox A; *TSPAN8* = tetraspanin 8; *LGR5* = leucine-rich repeat containing G protein-coupled receptor 5 ; *PRC1* = protein regulator of cytokinesis 1; *ZFAND6* = zinc finger; AN1-type domain 6; *FTO* = fat mass and obesity associated; *HNF1B* = HNF1 homeobox B; DUSP9 = dual specificity phosphatase 9; *SLC44A3* = solute carrier family 44; member 3; *F3* = coagulation factor III (thromboplastin; tissue factor); *RBM43* = RNA binding motif protein 43; *RND3* = Rho family GTPase 3; *GALNTL4* = polypeptide N- cetylgalactosaminyltransferase 18; *TMEM45B* = transmembrane protein 45B; *BARX2* = BARX homeobox 2 ; *CR2* = complement component (3d/Epstein Barr virus) receptor 2; *PCNXL2* = pecanex-like 2 (Drosophila); *PEX5L* = peroxisomal biogenesis factor 5-like; *PLS1* = plastin 1; *PSMD6* =proteasome (prosome; macropain) 26S subunit; non-ATPase; 6; *ZPLD1* = zona pellucida-like domain containing 1; *MAEA* = macrophage erythroblast attacher; *C6orf57* = chromosome 6 open reading frame 57; *KCNK16* = potassium channel; subfamily K; member 16; *ZFAND3* = zinc finger; AN1-type domain 3; *ACHE* = acetylcholinesterase (Yt blood group); *GLIS3* = GLIS family zinc finger 3; *PTPRD* = protein tyrosine phosphatase; receptor type; D; *KIF11* = kinesin family member 11; *TCERG1L* = transcription elongation regulator 1-like; *SPRY2* = sprouty homolog 2 (Drosophila); *C2CD4B* = C2 calcium-dependent domain containing 4B; *CMIP* = c-Maf inducing protein; *WWOX* = WW domain containing oxidoreductase; *SRR* = serine racemase; *LPIN2* = lipin 2 ; *HUNK* = hormonally up-regulated Neu-associated kinase ; *GRB14* = growth factor receptor-bound protein 14; *ST6GAL1* = ST6 beta-galactosamide alpha-2;6-sialyltranferase 1

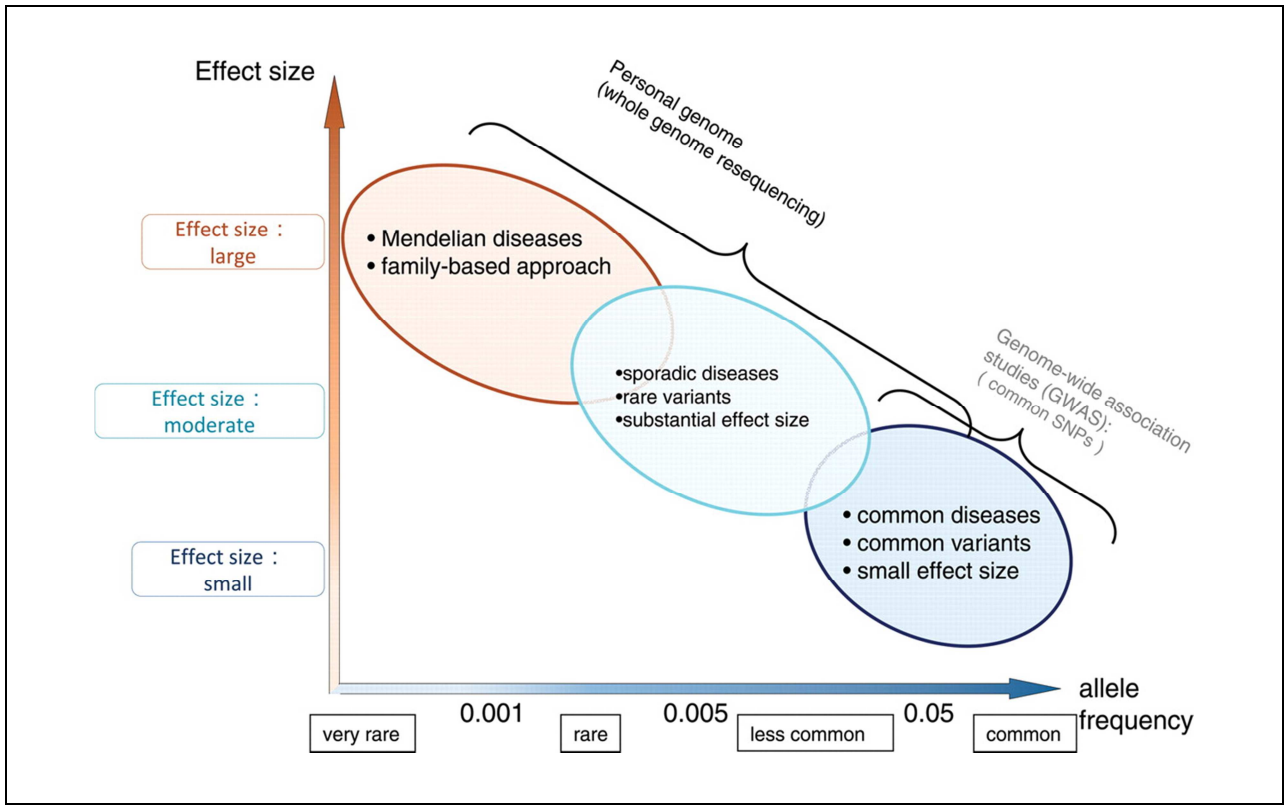
## 2.6 MISSING HERITABILITY OF GWAS-IDENTIFIED T2D ASSOCIATED SNPS

To date about 79 SNPs have been associated with T2D through GWASs that have been done among people of various ethnicities, as indicated in Table 2.4 (Sanghera & Blackett, 2012). However, there is still some missing heritability, which cannot be accounted for by the documented SNPs, as they have been reported to explain only 10% of the heritability of T2D among European populations (Voight et al., 2010).The missing heritability of GWAS-identified T2D variants has been an issue of great concern and several potential reasons for it have been suggested (Groop & Pociot, 2013). GWASs are not capable of identifying rare variants, which might have greater effects than the common variants that the GWAS methods are able to detect (Bush and Moore, 2012). Lack of clarity and failure to account for gene-gene and gene-environment interactions have also been postulated to be contributing to the missing heritability depicted by GWAS-derived variants (Sanghera & Blackett, 2012). These factors are discussed in greater detail in the following sections.

### 2.6.1 Rare gene variants and missing T2D heritability

Common variants that are detected via the GWAS approach are not the only mutations that are postulated to account for genetic susceptibility to complex diseases, as illustrated in Figure 2.3 below (Tsuji, 2010). Evidence from other genetic studies indicates that the genes that carry common variants with small effect sizes also harbour rare genetic variants with large effects on disease susceptibility (Zeggini *et al*., 2008). The GWAS approach is insufficiently powered and has a low resolution to be able to detect rare variants (Ahlqvist *et al*., 2012). Common variants of allelic frequencies greater than 5% and small effect sizes on disease susceptibility, as indicated by an odds ratio between 1.2 and 1.3, are usually discovered through the GWAS approach (McCarthy *et al*., 2008). It is also hypothesised that rare variants from many genes, when clustered together with common variants, might have a significant genetic effect, as indicated in the common disease-rare variant hypothesis, and could be better representative of complex disease genetic susceptibility (Petretto *et al*., 2007; Bodmer, 1999). The advent of next generation sequencing methodologies is expected to help discover the rare variants, which might help to decipher the missing heritability of T2D from GWAS (Groop & Pociot, 2013).

**Figure 2.3 Spectrum of disease-causing variants and their effect sizes.**




Adapted from Tsuji. (2010)

### 2.6.2 Copy number variations and missing T2D heritability

Copy number variations (CNV) are depicted as sub-microscopic deletions, duplications, inversions and translocations, which occur in small DNA segments (Lupski, 2007). Three CNV loci have been associated with diabetes (Craddock *et al*., 2010). CNVR2845.46 (HLA) and CNV7113.6 (Chr 17q) were associated with T1D and CNVR5583.1 (TSPAN8) with T2D (Craddock *et al*., 2010). CNVs are also suggested to account for the missing heritability of T2D (Sanghera & Blackett, 2012). However, very few CNV associations have been made with T2D, since the methodology is relatively new. However, the future generation of CNV maps is postulated to aid in the understanding of the missing heritability of T2D (Jakobsson *et al*., 2008; Craddock *et al*., 2010

### 2.6.3 Gene-gene interactions and missing heritability

The current statistical methods used in genetic analysis have been noted to possibly fail to account for the different effects of individual SNPs and their interactions by assuming

that all the SNPs have similar additive effects (Groop & Pociot, 2013). Epistasis is the interaction of genetic variants at two or more loci, resulting in a phenotypic outcome that is different from the additive effects of the individual loci (Churchill, 2013). GWASs have been reported to be insufficiently powered to enable assessment of epistasis in complex disease traits (Nair *et al.,* 2009). However, Evans *et al.,* (2006) simulated an approach that bypassed the power challenges of studying epistasis in GWASs through testing all pair-wise combinations of SNPs. It was noted that the power required for the evaluation of epistasis in GWAS was increased regardless of multiple testing when no marginal effects were included (Evans *et al*., 2006). However, studies of epistasis in GWASs using the approach of Evans *et al*. (2006) are limited and more novel methodologies are required to unravel the missing heritability of T2D due to epistasis (Groop & Pociot, 2013).

### 2.6.4 Gene-environment interactions and missing heritability

The recent surge in the prevalence of T2D is suggested to be largely a result of the increase in the adoption of Western lifestyles among people with prior T2D genetic susceptibility (Franks *et al.*, 2013). However, in GWAS, limited interactions of SNPs with environmental factors and the effect of such interactions on diabetes or its related traits have been detected (Sanghera & Blacket, 2012). Inaccurate measurements of environmental factors such as diet and exercise have been indicated as key limitations leading to failure to detect interactions of genetic variants with environmental factors in GWAS (Franks, 2011). However, in other studies the interaction of GWAS-derived SNPs with environmental factors such as diet has been determined (Qi *et al.,* 2009). Adoption of the Western dietary pattern among men who had high genetic risk scores of GWAS-derived variants was associated with an odds ratio of 2.06 for developing T2D, which was higher than the 1.2-1.3 odds ratios reported for GWAS variants alone (Qi *et al.,* 2009). The failure to detect gene-environment interactions accurately is thus postulated to contribute to the T2D missing heritability of the GWAS approach (Sanghera & Blackett, 2012).

### 2.6.5 Epigenetics and missing heritability of T2D

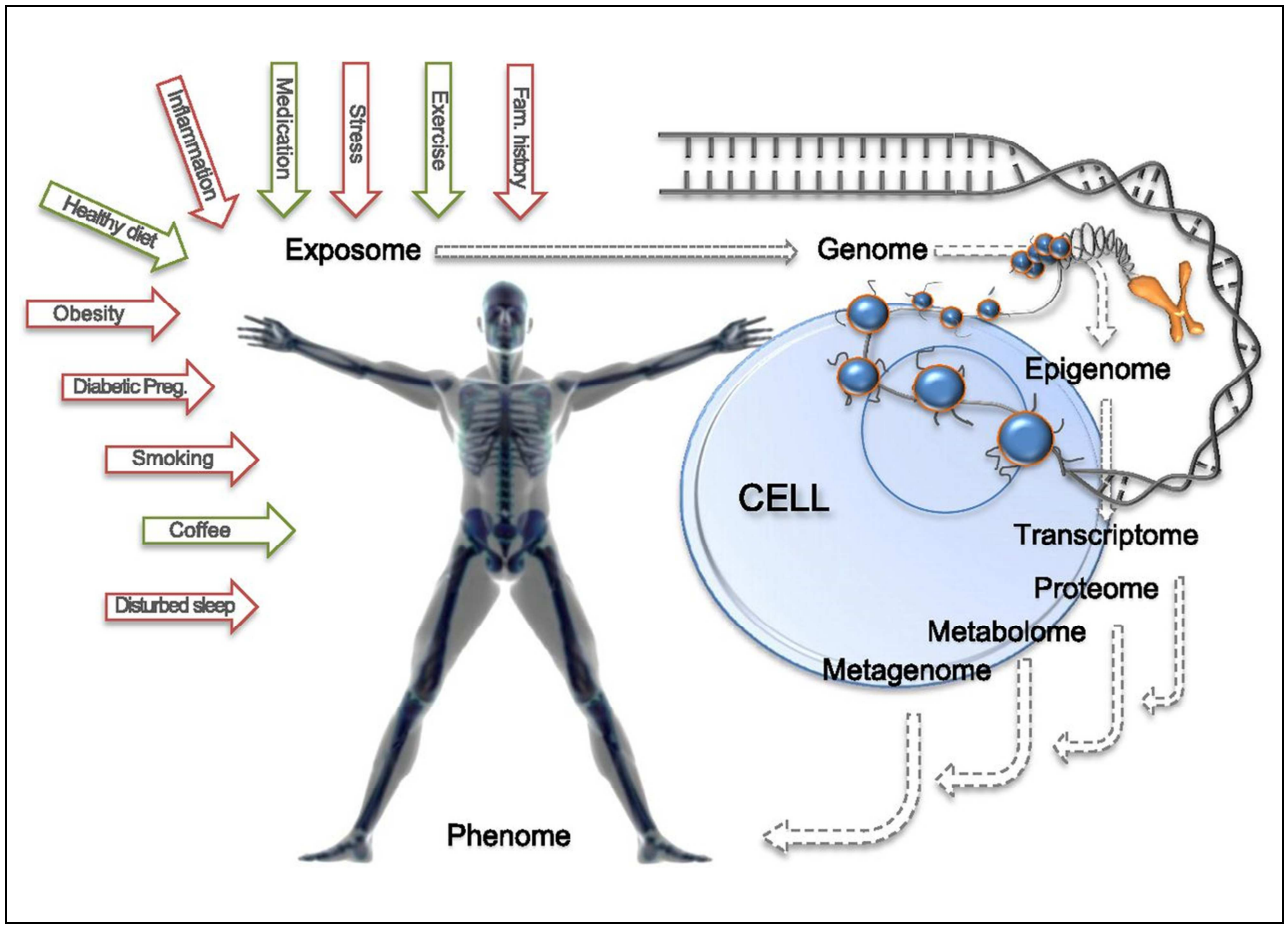
Changes in the epigenome and not just variations in DNA sequence have also been reported to increase disease susceptibility through altering the expression of the affected genes (Groop & Pociot, 2013)**.** Epigenetic changes include the methylation of DNA, post-translational modification of histones and activation of micro-ribonucleic acid (Chong & Whitelaw, 2004). The epigenetic changes have been noted to be stable and heritable across cell division in the human body (Liu *et al*., 2008). Through the "metabolic memory" hypothesis it is suggested that glucose can induce histone modifications in endothelial cells, which can be memorised for years to come (Brasacchio *et al*., 2009). Epidemiological evidence of the link between early nutrition, foetal growth and development of chronic diseases such as T2D has been postulated to be a result of epigenetic mechanisms (Barker, 2007). However, GWASs have evaluated only changes in DNA structure; the epigenetic factors were not considered and may thus be part of the missing heritability of T2D (Groop & Pociot, 2013)**.**

### 2.6.6 Systems biology and missing T2D heritability

There is now growing concern that GWASs provide limited insight into the genetic aetiology of complex diseases (Groop & Pociot, 2013). Much focus in GWAS has been on the genome; however, there are other critical factors, which have been ignored, that play pivotal roles in disease pathogenesis. The genes are transcribed and translated into proteins and there are other processes and pathways besides the genome that can alter the proteins, leading to the development of a disease (Groop & Pociot, 2013). The systems biology approach, which involves the analysis and interpretation of the interrelationships of the varied elements in the human system, including metabolites, processes, networks and pathways, is deemed to shed more light on the role played by the GWAS-derived variants in the context of the other bodily functions and processes to help unravel the genetic susceptibility of complex diseases (Mesarovic *et al*., 2004; Suhre *et al*., 2011). Network and pathway-based approaches have been noted to aid in identifying disease genes (Bergholdt *et al*., 2012). A recent study identified novel genes associated with T2D, which could account for 25% of the variance in HbA1C by using expression differences and co-expression profiles of reported genes and protein–protein interaction analyses of human pancreatic islets (Taneera *et al.,* 2012). Therefore a

systems-based approach based on understanding aetiological mechanisms that integrate the GWAS-derived variants is hypothesised to help unravel the missing heritability of T2D, as illustrated in Figure 2.4 (Bergholdt *et al*., 2012; Rossin *et al*., 2011).

**Figure 2.4 The future of systems biology approach for studying T2D**




Adopted from Franks *et al*. (2013)

## 2.7 TYPE 2 DIABETES GENETIC STUDIES IN AFRICA

Africa has been reported to be lagging behind in deciphering the genetic susceptibility of various diseases (Ramsay, 2012). Low levels of funding; poor infrastructure and limited numbers of trained scientists are some of the reasons for the slow take-off of GWASs in Africa (Ramsay, 2012). The H3 Africa initiative, which is funded by the National Institutes of Health (USA) and Wellcome Trust (UK) was recently launched to enhance the drive to elucidate the genetic susceptibility to selected diseases, which include T2D, in Africa (Ramsay, 2012). Africa is reported to harbour the greatest amount of genetic

diversity among humans when compared to other continents (Schuster *et al*., 2010). The sequencing of the exomes of three San groups who had different languages and lived in different areas revealed large amounts of genetic diversity that was noted to be greater than that between Europeans and Asians (Schuster *et al*., 2010). Thus, it is envisaged that the inclusion of the Yoruba and Kenyan populations in the HapMap and 1000 Genomes Project will be under-representative of the genetic variation present in Africa (Ramsay, 2012; May *et al*., 2013). The genetic variation in Africa and how it relates to disease susceptibility remains insufficiently studied (Ramsay, 2012).

The genome-wide linkage analysis approach has been extensively used to assess T2D genetic susceptibility studies in African countries, which include Nigeria and Ghana (Rotimi *et al*., 2004; Chen *et al*., 2007). However, in view of the failure of the linkage analysis approach to aid in deciphering the genetic architecture compared to GWASs, there is an urgent need to evaluate GWAS-derived variants in the African context as well. Although the GWAS variants derived from the European context have been noted to account for only 10% of the heritability of T2D, little is reported with regard to their effects in African populations (Sanghera & Blackett, 2012). Genetic diversity has been noted to exist among people of different ethnicities and this has led to the realisation that GWAS findings from one population group are not suitable to predict genetic risk in other populations (Kivisild *et al*., 1999; Sanghera & Blackett, 2012). However, lessons can be learnt on how to assess the genetic variation in Africa from other non-European ethnicities that have reported about the shared variants associated with T2D among people of different ethnicities.

### 2.7.1 Shared type 2 diabetes gene variants among different ethnicities

There is growing evidence that regardless of the differences in genetic diversity, some common variants associated with T2D are shared among people of various ethnicities (Cauchi *et al*., 2012). In a large multi-ethnic study, comprising European Americans, African Americans, Latinos, Japanese Americans and Native Hawaiians, 14 variants were noted to have similar association patterns among all the ethnicities that were being studied (Waters *et al*., 2010). Recently, 40 European GWAS-derived T2D-related variants were evaluated among North African Arabs and 13 variants were noted to be similarly associated between North African Arabs and Europeans (Cauchi *et al*., 2012).

It is therefore plausible that there might be T2D variants that might be shared among black South Africans and other ethnicities. However, a majority of the common variants that have been associated with T2D in other ethnicities have not yet been evaluated among black South Africans.

### 2.7.2 GWAS-determined common variants among black South Africans and T2D genetic susceptibility

Limited evidence exists regarding the association of GWAS-derived variants with T2D susceptibility in black South Africans. Pirie *et al.* (2010) evaluated variants in PPARG, KCNJ11, TCF7L2, FTO and HHEX genes among South Africans of Zulu descent and discovered only rs7903146 in the TCF7L2 gene to be significantly associated with T2D, with an odds ratio of 1.84. Many of the common variants associated with T2D in other ethnicities are yet to be evaluated in the black South African setting. Recently, it has been proven that there is a great amount of genetic diversity between black South Africans and the Yoruba people used in the HapMap project to capture genetic variation among Africans (May *et al*., 2013). Therefore, the under-representation of the black South African genetic variation in HapMap, which was used as a proxy for designing GWAS chips, denotes that fewer private or population-specific variants associated with T2D susceptibility will be detected when the SNPs from GWAS are evaluated in the South African context (May *et al*., 2013). However, Teo *et al.* (2010) argue that the low LD present among Africans has the advantage of enabling the detection of causal variants, compared to European or Asian populations, which have high LD (Teo *et al*., 2010). Therefore evaluating GWAS-derived genetic variants associated with T2D in other ethnicities is pivotal in the black South African population as an initial strategy for deciphering the genetic architecture of this disease, which can also lead to the detection of causal variants.

## 2.8 SUMMARY

T2D is a multifactorial disease whose aetiology comprises genetic and environmental factors that interact in a complex manner. Evidence exists that T2D has a strong genetic component, which is yet to be fully elucidated. Various genetic approaches were adopted to decipher the genetic landscape of T2D. Of these, the GWAS

methodology has greatly improved current knowledge regarding T2D genetic susceptibility, with 79 common genetic variants being discovered through this approach. However, these current variants have been noted to account for only about 10% of the T2D genetic susceptibility. Much is still required to unravel the missing T2D heritability, which has not been detected by the GWAS approach. It is hypothesised that the common variants determined from GWAS and rare variants that will be detected by genome sequencing, together with a systems biology approach, will help to unravel T2D missing heritability in future. African countries are lagging behind in deciphering T2D genetic susceptibility. Although the reported GWAS variants have been noted to account for only 10% of T2D heritability in European populations, a large number of them have not been evaluated in the black South African context. Evidence of common T2D variants that are shared among different ethnicities is a motivation to evaluate the common genetic variants that were discovered among other ethnicities in the black South African population. Evaluation of the common genetic variants from other ethnicities is a pivotal initial step in unravelling the T2D genetic susceptibility of black South Africans.

# CHAPTER 3

# METHODOLOGY

The current chapter discusses the techniques and methods undertaken to attain the objectives of this study, as outlined in Chapter 1. Section 3.1 provides information about the ethical aspects of the study. The characteristics of the study sample are discussed in Section 3.2, while Section 3.3 discusses how the sample size was determined. The calculation of BMI and the manner in which the questionnaires about the dietary habits of the research subjects were dealt with are outlined in Section 3.5 and Section 3.6 respectively. Section 3.7 discusses the manner in which the SNPs for this study were selected and genotyped and the way in which the results were validated. Section 3.8 elaborates on the process by which the descriptive statistics for the research subjects were determined, using the SPSS version 21 software program; the genetic associations were determined using the gPlink software version 2.050.

## 3.1 STUDY DESIGN

This study was a nested case-control study within the South African arm of the Prospective Urban Rural Epidemiology (PURE) study, of which the baseline data was collected in 2005. Matching of cases and controls was performed on an individual basis for age and sex.

This study was performed according to the principles of the Declaration of Helsinki, 2013, and all participants were requested to give written informed consent before participating in the study. The PURE study was granted ethical approval by the Ethics Committee of the North-West University, Potchefstroom Campus, with ethics number NWU-00016-10-A1 and this study was later approved with ethics number NWU-00016-10-S1 by the same ethics committee.

## 3.2 STUDY POPULATION

The PURE study is a large-scale epidemiological study, which has recruited approximately 140 000 individuals residing in more than 600 communities in 17 low, middle, and high income countries around the world. Individual data collection consists of the determination of medical history, lifestyle behaviour (physical activity and dietary profile), blood collection (for both genetic and biochemical analyses), an electrocardiogram, and anthropometric measures. In addition, detailed information is collected with respect to four specific environmental factors of importance, which include the level of urbanisation, nutrition and associated food policy, psychosocial/socioeconomic factors and tobacco usage (Teo *et al*., 2009). The study was begun in 2005 and entails five-year follow-up intervals up to 2015, with the study ending in 2017 (Teo *et al*., 2009).

In the larger PURE study, males and females between the ages of 35 and 60 years who were healthy, as indicated by not using any medication for chronic disease and not having a diagnosed medical condition/disease, were recruited. The participants were recruited from urban sections (Ikageng and Sonderwater extension 7 and 11) and in rural areas (Ganyesa and Tklagameng). Residents were defined as people who slept most times per week at a specific household and who considered it as their habitation on a long-term basis. Renters, boarders and servants were not considered household members.

The study participants for this specific study were recruited from the Profiles of Resistance to Insulin in Multiple Ethnicities and Regions study, which consisted of a random sample of 443 participants from the PURE study (Kotze, 2010). The majority of the research participants identified themselves as Tswana-speaking. For this study, the cases were 180 research subjects who had IGT as diagnosed by having plasma glucose levels above 7.8 mmol.$L^{-1}$ after a two-hour OGTT and they were matched for sex and age with 180 controls that had normal glucose levels less than 7.8 mmol.$L^{-1}$ following a two-hour OGTT (ADA, 2012).

## 3.3 SAMPLE SIZE CALCULATION

The QUANTO Version 1.2 (2007) genetics power calculator was used for the sample size calculation. A power of 86% to detect SNPs of allele frequencies of more than 20% with an odds ratio of 1.65 or greater at a significance level set at 0.05 could be attained using a sample size of 180 cases. Although the sample size was low compared to the size of about 1 000 recommended for replication studies of GWAS, it was deemed suitable in view of local studies of similar size, which had been able to replicate T2D-associated variants among black South Africans (Pirie *et al*., 2010; Lewis & Knight, 2012).

## 3.4 ORAL GLUCOSE TOLERANCE TEST

The OGTT was performed by appropriately trained research nurses on the subjects who were included in the investigation. The subjects were required to have no acute or chronic infections and to have fasted for at least 10 hours before the OGTT (Kotze, 2010). Normal fasting blood glucose levels were first determined before the OGTT was administered to screen for undiagnosed T2D patients, by taking a finger prick measurement with the MediSens® Optium™[1] Sensor glucose apparatus (Kotze, 2010). Intra-venous catheterisation was performed to enable the collection of venous blood. The participants were given a 75 gram glucose load dissolved in 300 mL of water to ingest. The blood samples were collected before the consumption of the glucose load (at 0 min) and post-prandially at 30, 60, 90 and 120 minutes for glucose level measurements. Both capillary and plasma blood glucose levels were measured via a handheld MediSens® Optium™ Sensor glucose apparatus, together with MediSens® Optium™ Plus blood glucose test strips by the research team members (Kotze, 2010).

## 3.5 ANTHROPOMETRY

The anthropometric measurements were determined according to the guidelines adopted at the Arlie Conference (Lohman *et al*., 1988). The weight of the subjects was measured by weighing participants in minimal clothing, barefoot and with arms hanging

[1] MediSens® Optium™ is a registered trademark of Abbott Diabetes Care, Doncaster, Australia

loosely at their sides. Measurements were performed in duplicate by the PURE research team members using a portable electronic scale (Precision Health Scale, A&D Company, Tokyo, Japan), after which the mean was recorded (Lohman *et al*., 1988). If there was a difference of greater than 0.1 kg between the duplicate measures, a third measurement was taken.

The heights of the subjects were determined by the PURE study research team members using a stadiometer (IP 1465, Invicta, and London, UK). The subjects were required to be barefoot, with their hands hanging loosely at their sides and their heads held in the Frankfort plane (represented by a line between the margin of the orbit of the eye and the tragion) during the height measurements. The height was taken twice to the nearest 0.1 cm and the mean was used for further statistical analysis (Lohman *et al*., 1988). The BMI of the participants was computed using the formula indicated in Equation 3.1 below.

**Equation 3.1 Formula for calculating BMI**

$$\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height (m}^2\text{)}}$$

BMI= body mass index; kg = kilograms; m = metres

## 3.6 QUESTIONNAIRES

The subjects completed specific questionnaires in the language of their choice with the assistance of the researchers and trained fieldworkers. The questionnaires were used to capture information about age, gender, medical history, residence area (urban or rural) and tobacco use. The captured information from the questionnaires was entered into the Microsoft EXCEL®[2] spreadsheet programme and exported to the Statistica®[3] version 10 program for statistical analysis.

The trained fieldworkers captured the dietary intake of the participants using a quantitative food frequency questionnaire (QFFQ). Food portion books, which had been

[2] EXCEL® is a registered trademark of Microsoft Office. Redmond, WA, USA
[3] Statistica® is a registered trademark of Statsoft Inc., Tulsa, Oklahoma, USA

designed, standardised and validated for use in the North West Province of South Africa, were used to aid the estimation of portion sizes using household measures. Earlier studies by MacIntyre *et al*. (2001a) and MacIntyre *et al*. (2001b) had initially validated the QFFQ for use in South Africa. The dietary intake data was analysed and nutrient intakes which comprised of carbohydrate and fat intake were computed using the Foodfinder3®[4] program at the Medical Research Council of South Africa.

## 3.7 GENETIC ANALYSIS

The genetic analyses were performed in a sequence of several steps. Firstly, a literature search was conducted to identify the previously documented SNPs associated with T2D in other ethnicities. Secondly, the identified SNPs were evaluated to determine the probability of successfully genotyping them with the BeadXpress®[5] technology. Thirdly, kits were designed for the SNPs, which had a high probability of being successfully genotyped by Illumina and the DNA samples were normalised according to criteria required for genotyping by means of the BeadXpress® technology. The samples were then genotyped. The steps described above, which were undertaken during the genetic analysis, are elaborated on in detail in the paragraphs outlined below.

### 3.7.1 Literature search of SNPs associated with T2D

A search strategy using the keywords “genetics, type 2 diabetes, SNPs, variants, GWAS” was used to look for the reported common variants associated with T2D, using selected databases of research articles. PubMed, EBSCOhost and Science Direct were the databases used to search for review articles that had reported on the common variants associated with T2D through GWAS studies up to February 2013. The review articles were analysed to determine those that contained a greater representation of the SNPs associated with T2D risk in multiple ethnicities. The SNPs presented in the Table 2.4 were selected for this study from the identified review article.

[4] The dietary data was computerised and analysed by Ria Laubscher, at the Medical Research Council, South Africa using *Foodfinder3*® a registered trademark of the Medical Research Council, Tygerberg

[5] BeadXpress® is a registered trademark of Illumina® Inc., San Diego, CA, USA

### 3.7.2 Pre-assessment of SNPs for genotyping

The SNPs were sent to Illumina®[6] to be pre-assessed for suitability for genotyping on the BeadXpress® genotyping platform. The Assay Design Tool (ADT) was used to assess the chances of successfully genotyping the selected SNPs. An SNP score list was generated, which indicated the position of the variants, minor allele frequencies and metrics that were used to determine the chances of successfully genotyping the SNPs, as discussed in Section 4.3. The kits were designed by Illumina for the SNPs that had acceptable metrics for successful genotyping on the BeadXpress® platform as determined by the ADT. The completed kits were shipped to the Animal Breeding and Genomics Centre at the Wageningen University, where genotyping was performed using the BeadXpress® platform.

### 3.7.3 DNA isolation

The genomic DNA was isolated from citrate-treated buffy coat using the modified protocol of the Qiagen® Flexigene™[7] DNA extraction kit (catalogue number 51206). One mL of buffy coat was mixed with 2.5 mL of lysis buffer (FG1) by inverting the mixture in a centrifuge tube five times. The mixture of FG1 buffer and buffy coat was centrifuged for five minutes at a gravitational force of 2 000 x g. The supernatant was discarded and the tube inverted on a clean sheet of absorbent paper for two minutes while taking care that the pellet remained in the tube. A solution consisting of 0.5 mL denaturation buffer (FG2) and 5 µL of Qiagen® protease was added to the tube that contained the pellet. The tube was incubated at 65°C for 10 minutes. After incubation, 0.5 mL of isopropanol was added to the tube and mixed by inversion until the DNA was visible as a clump. Centrifugation of the sample was undertaken for three minutes. The supernatant was discarded and 0.5 mL of a 70% ethanol solution was added to the DNA. The DNA pellet was washed by inverting the tube several times. Centrifugation was repeated and the ethanol was aspirated. Two hundred microlitres of FG3 buffer was added after low-speed vortexing had been performed to dissolve the DNA and the solution was incubated at 65° C for one hour. The NanoDrop®[8] spectrophotometer (ND-1000) was used to determine the purity and yield of DNA at an absorbance of 260 nm

[6] Illumina is a registered trademark of Illumina, Inc in San Diego, CA, USA
[7] FlexiGene™ is a trademark of the QIAGEN Pty. Ltd., Australia
[8] NanoDrop is a trademark of Thermo Fisher Scientific, Wilmington, USA

and 280 nm. The assessment of the DNA purity and the principles of this process are explained in Section 3.7.4. The stock solution was stored at minus 20°C and thawed at room temperature for use as elaborated in Section 3.7.4.

### 3.7.4 Normalisation of DNA

The DNA stock solution, which had previously been extracted by the Nutrigenetics research team at the North-West University, was normalised to a concentration between 45 and 55 ng.µL$^{-1}$, which is the concentration required for analysis via the BeadXpress® apparatus. Double-distilled water was used to dilute the DNA samples to the required 50 ng. µL$^{-1}$ in a 20 µL volume using the formula indicated in Equation 3.2.

**Equation 3.2 Normalisation of DNA to 50 ng.µL$^{-1}$**

$$C_1V_1 = C_2V_2$$

$$D = V_2 - V_1$$

$C_1$ = mean of the two DNA concentrations of the stock solution in ng.µL$^{-1}$; $V_1$ = volume of DNA from the stock DNA solution in µL required; $C_2$ = [DNA] in the working dilution (50 ng. µL$^{-1}$); D = volume of double-distilled water to be added to the working dilution in µL; $V_2$ = final DNA volume (20 µL)

The working dilution of DNA was assessed for purity and yield using the NanoDrop® spectrophotometer. The A260/A280 and the A260/A230 ratios were used to assess for contamination of the DNA with proteins or ethanol. The DNA was considered pure if it had an A260/A280 absorbance ratio above 1.8 and an A260/A230 absorbance ratio within the range of 1.8-2.2. The normalised DNA samples were plated onto 96-well plates and shipped to the Animal Breeding and Genomics Centre at the Wageningen University, where researchers genotyped them through Illumina® BeadXpress® technology, using the kits that had been designed by Illumina after the pre-assessment, which is described in Section 3.7.2.

### 3.7.5 Genotyping with Illumina BeadXpress® Technology

The GoldenGate®[9] Assay with VeraCode®[10] technology was used for genotyping the selected SNPs using the BeadXpress® platform. The work flow for the genotyping was split over two days. On the first day the DNA was initially activated to enable it to bind

[9] GoldenGate® is a registered trademark of Illumina® Inc., San Diego, CA, USA
[10] VeraCode® is a registered trademark of Illumina® Inc., San Diego, CA, USA

to paramagnetic particles. The activation step was robust and required a minimum of 250 ng of genomic DNA. The activated DNA was combined with paramagnetic particles, assay oligonucleotides and hybridisation buffer to enable hybridisation. Three oligonucleotides, which consisted of two allele-specific oligonucleotides (ASO) and locus-specific oligonucleotides (LSO), were designed for each SNP locus and included in the assay mix. The LSO contained a unique address, which enabled it to bind to a particular VeraCode® micro-bead type. After the hybridisation step, excess and mishybridised oligonucleotides were removed through several washing steps. Extension of the appropriate ASO and ligation of the extended product to the LSO joined the information about the genotype present at the SNP site to the address sequence on the LSO. The ligation products were used as polymerase chain reaction (PCR) templates for the universal PCR primers P1, P2, and P3. The primers P1 and P2 were labelled with Cy3 and Cy5 respectively and the PCR cycle was performed.

On the second day, the dye-labelled single-stranded PCR products were hybridised with the complementary bead types by using the unique address sequences. The binding of the GoldenGate® Assay products on to the VeraCode® beads enabled the separation of the assay products, allowing the individual SNP genotypes to be read on the BeadXpress® reader. The BeadXpress® reader identified the microbead code and the fluorescent signals by using a laser beam, which penetrated the VeraCode® microbead. The amplification product for each SNP displayed a signal in either the Cy3 or Cy5 channels if the participant harboured either of the homozygous genotypes, while those harbouring the heterozygous genotype presented with a signal in both channels. Finally, the Illumina GenomeStudio® Genotyping Module v1.0 program was used to analyse the data from the BeadXpress® reader to enable genotype clustering and the calling of the genotypes. The GenomeStudio® results and the controls that were used in the genotyping process are discussed in more detail in Section 4.3.4.

### 3.7.6 Validation of the BeadXpress® genotyping

Samples were randomly selected from the clusters of selected SNPs for validation through automated Sanger sequencing. Primers were first designed for the selected

SNPs using the Beacon Designer™[11] 8.10, a program used to design primers for PCR amplification. The amplification of the DNA regions of the samples was performed and the PCR products were sent for sequencing at the Central Analytical Facility at Stellenbosch University, as described in the following sections.

### 3.7.7 Primer design

The DNA sequences of the regions that harboured the SNPs of interest were entered into the Beacon Designer™[12] 8.10, a program used to design PCR primers. The forward and reverse primer pairs of the selected SNPs were suggested and ranked by the program. The primer pairs were rated and ranked according to characteristics that could enable successful amplification. These primer properties included primer length between 18 and 25 bases, having an annealing temperature difference of less than 5°C between the primer pairs, a GC content of 45% to 60% and a small possibility to no possibility of forming hairpins or self/cross-dimers. The primers with the highest rates of successful amplification were selected. The primers were evaluated through the NCBI Primer Blast tool to assess if they also annealed to other sections of the human genome besides those of the selected SNPs. Primer sets that were specific to only the DNA regions of interest were selected and submitted to be manufactured by Whitehead Scientific. The primers indicated in Table 3.1 were used for the amplification process.

**Table 3.1 Characteristics of the PCR primers**

| Primer | Primer Sequence (5'-3') | Annealing Temperature | Product size (bp) |
|---|---|---|---|
| rs 896854F | CCTTAACATTTAGATCCTTT | 58.3°C | 565 |
| rs 896854R | CCAGTTCAGAATTACACT | | |
| rs 6467136F | CTCATTACCTCATATCATA | 54.7°C | 798 |
| rs 6467136R | TTATCAATCAATGGACAT | | |
| rs 7630877F | TGAGTTGAATATTATTGTTG | 56.5°C | 730 |
| rs 7630877R | GTAGAATTAAGCATTGTG | | |

bp = base pairs; T = thymine; A = adenine; C = cytosine; G = guanine; F = forward; R = reverse; °C = degree centigrade

[11] Beacon Designer™ is a trademark of PREMIER Biosoft International, USA
[12] Beacon Designer™ is a trademark of PREMIER Biosoft International, USA

### 3.7.8 Amplification conditions

The Biorad CFX96 Touch™[13] Real time PCR machine was used to perform the PCR reactions. The PCR reaction mix consisted of 8.875 µL of PCR grade water, 5 µL of buffer, 45 nmol of magnesium chloride, 10 nmol dNTPs, 1.25 U of polymerase enzyme, 0.63 pmol each of the forward and reverse primer, and 50ng of DNA to make up a volume of 25 µL.

The PCR amplification was achieved using the following cycle conditions: The initial denaturation at 95°C for three minutes was followed by 40 cycles that consisted of denaturation at 95°C for 30 seconds, annealing at the specific annealing temperatures ($T_a$) for 40 seconds and extension at 72°C for 45 seconds. The final cycling conditions were elongation at 72°C for five minutes and a hold step at 4°C indefinitely.

The PCR reaction was evaluated via electrophoresis using the FlashGel™[14] system for eight minutes at 150 volts. An image of the separated fragments was captured by the FlashGel™ camera and their sizes determined by comparing them to the FlashGel™ (100-4000 bp) DNA molecular weight marker manufactured by Lonza. The amplification was regarded as successful when the fragment size matched the expected PCR product size.

### 3.7.9 Sequencing

The PCR products were submitted for sequencing to the Central Analytical Facility at Stellenbosch University. The sequencing was performed according to the principles of Sanger sequencing (Sanger *et al*., 1977). Initially, the PCR products were purified using NucleoFast®[15] 96 PCR Clean-Up kits. The PCR clean-up was undertaken to separate the amplified products from contaminants, which included primers, deoxyribonucleotide triphosphate (dNTPs) and salts. The NucleoFast® technology uses ultrafiltration principles to separate the contaminants from the PCR product. During the clean-up process, the amplified product is passed through a membrane under pressure in a vacuum to separate the contaminants from the PCR product, leaving it retained on

[13] BioRad CFX96 Touch™ is a trademark of Bio-Rad laboritories, Inc., California, USA
[14] FlashGel™ is a trademark of Lonza, Rockland, USA.
[15] NucleoFast® is a registered trademark of Machery-Nagel, Germany

the membrane. The PCR product is then recovered by incubating the membrane for one to three minutes with nuclease-free water or a low-salt buffer.

The sequencing was conducted after the PCR clean-up. The ABI Prism®[16], BIGDye®[17] Terminator version 3 Ready Reaction Cycle Sequencing Kit supplied by Applied Biosystems was used for the sequencing reactions. The automated sequencing method that was used is based on the principle of chain termination sequencing (Sanger *et al*., 1977). Initially, the DNA is denatured into a single strand. Primers specific to the DNA fragment to be sequenced, DNA polymerase, dNTPs and fluorescently labelled 2', 3'-dideoxyribonucleotide triphosphates (ddNTPs) are added to the mixture. A new strand of DNA begins to be formed when the DNA polymerase binds to the primer and starts to incorporate the dNTPs. However, when the DNA polymerase incorporates the fluorescently labelled ddNTPs the elongation reaction terminates. A large number of DNA fragments of different lengths ending in fluorescently labelled nucleotide bases are made, as the process of sequencing is repeated several times.

At the end of the sequencing reaction, the product was purified using a modification of the Princeton Separations post-sequencing kit, which is used to remove unincorporated fluorescently labelled ddNTPs. This purifying kit uses perforated beads to retain the ddNTPs and dNTPs while allowing the larger fragments to pass through. Following this, capillary electrophoresis was performed on an ABI3120xl genetic analyser according to standard protocols to separate the purified product into the differently sized fragments that were individually detected as they were excited by a laser beam, while they traversed the gel. The sequence electropherograms of the samples were analysed using the CLC Main Workbench 6.8.4 software of the Qiagen Company. The forward and reverse sequences for the selected SNPs rs896854, rs6467136, and rs7630877 were aligned with reference to sequences retrieved from dbSNP for the respective SNPs to identify if any alterations were present.

[16] ABI Prism® is a registered trademark of Applied Biosystems, California, USA
[17] BIGDye® is a registered trademark of Applied Biosystems, California, USA

## 3.8 STATISTICAL ANALYSIS

The descriptive statistics of the study participants were analysed using the statistical package for social scientists (SPSS) version 21 software program, whereas the statistical genetic analysis was performed using the gPLINK software version 2.050. Dr Suria Ellis, a statistician at the statistical department of the North-West University, Potchefstroom Campus, assisted the researcher with the statistical analysis for this study.

### 3.8.1 Descriptive statistics

The mean and standard deviations of the dietary, demographic and two-hour blood glucose values of the cases and controls were determined. The Q-Q plots were used to assess the normality of the continuous variables in this study to determine if they were suitable to be analysed by parametric tests. The central limit theorem was also used to consider whether the continuous variables in this study were fit for the use of the parametric tests. The central limit theorem, which implies that for a sample size greater than 30 the sample mean approaches normality, was adopted in view of the sample size of this study, which was much greater than 30 (Moore & McCabe, 2003). Thus, in view of the central limit theorem, it could be deemed that the continuous variables in this study could be treated as fit for parametric tests that require normally distributed variables. Statistical tests were undertaken to evaluate if there were differences in the characteristics of the cases and controls that would confound the genetic association tests. The Student t test was used to compare age and dietary intake parameters (which were continuous variables) between the cases and controls. The categorical variables of the cases and controls, which included sex and urbanisation, were compared using the chi-square test. A p-value of 0.05 was considered statistically significant.

### 3.8.2 Calculation of the distribution patterns of the selected SNPs

The Hardy-Weinberg equilibrium (HWE) status of the selected SNPs was determined in order to evaluate the distribution patterns of the selected SNPs. The observed genotypic frequencies of the selected SNPs were evaluated to see how well they

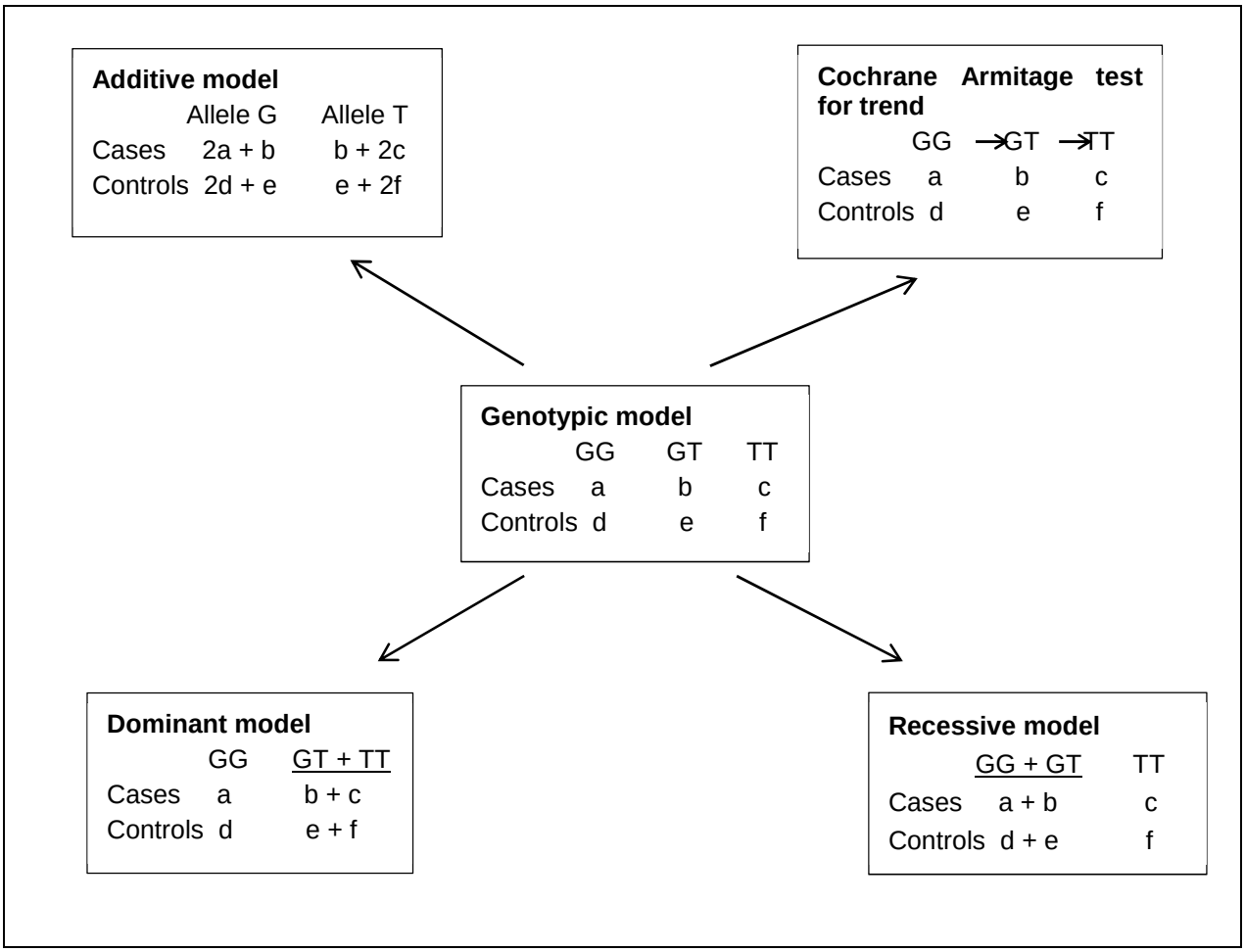
adhered to the assumptions of the HWE model. The HWE model assumes that the genotype frequencies are a result of random mating and are not influenced by the effects of genetic drift, mutation, migration, and natural selection. The observed genotypic frequencies were compared with the expected genotypic frequencies using the chi-square statistical method. A statistically significant difference between the observed and expected genotype frequencies, indicated by a p-value of less than 0.05, was used to denote that the SNP did not adhere to the assumptions of HWE. HWE is also an indirect index for genotyping errors. It has been reported that assays that have many SNPs that do not adhere to HWE are indicative of poor genotyping (Hosking *et al*., 2004). Thus the HWE was also used to evaluate SNPs that were not well genotyped.

### 3.8.3 Testing for genetic association of SNPs with IGT

The specific underlying models of inheritance of the selected SNPs were not reported; therefore all SNPs were evaluated using the standard models of genetic association. The additive, recessive, dominant, Cochrane Armitage test for trend (CATT) and general genotypic models were used to evaluate the association patterns of the selected SNPs. The genetic models are regarded as standard tests for disease penetrance (Clarke *et al*., 2011). The additive model evaluated the manner in which the differences in the order of increase in the number of minor alleles that were carried among the cases and controls were associated with T2D risk. The additive model, unlike other models of inheritance, evaluates the differences in disease association in reference to the alleles, not genotypes, as indicated in Figure 3.1 below (Lewis & Knight, 2012). The recessive genetic model was used to assess associations of the SNPs with IGT by comparing the genotypic distributions of the homozygous carriers of the minor allele with the carriers of the major allele among the cases and controls. The dominant model evaluated the association of the SNPs with IGT by evaluating the differences in the genotypic distributions of the carriers and non-carriers of the minor allele among the cases and controls, as illustrated in Figure 3.1. The general genotypic model assessed the association of IGT with the SNPs by comparing the differences in the genotypic distributions of the cases and controls without making assumptions about the manner in which the alleles were related to each other, as indicated in Figure 3.1. The CATT evaluated the manner in which the differences in the genotypic distributions

among the cases and controls, according to the increase in the number of the minor alleles carried by the participants, were associated with IGT, as indicated in Figure 3.1 below.

**Figure 3.1 Genetic models of disease penetrance**




Adopted from Lewis & Knight, (2012) T = thymine; A = adenine; C = cytosine; G = guanine

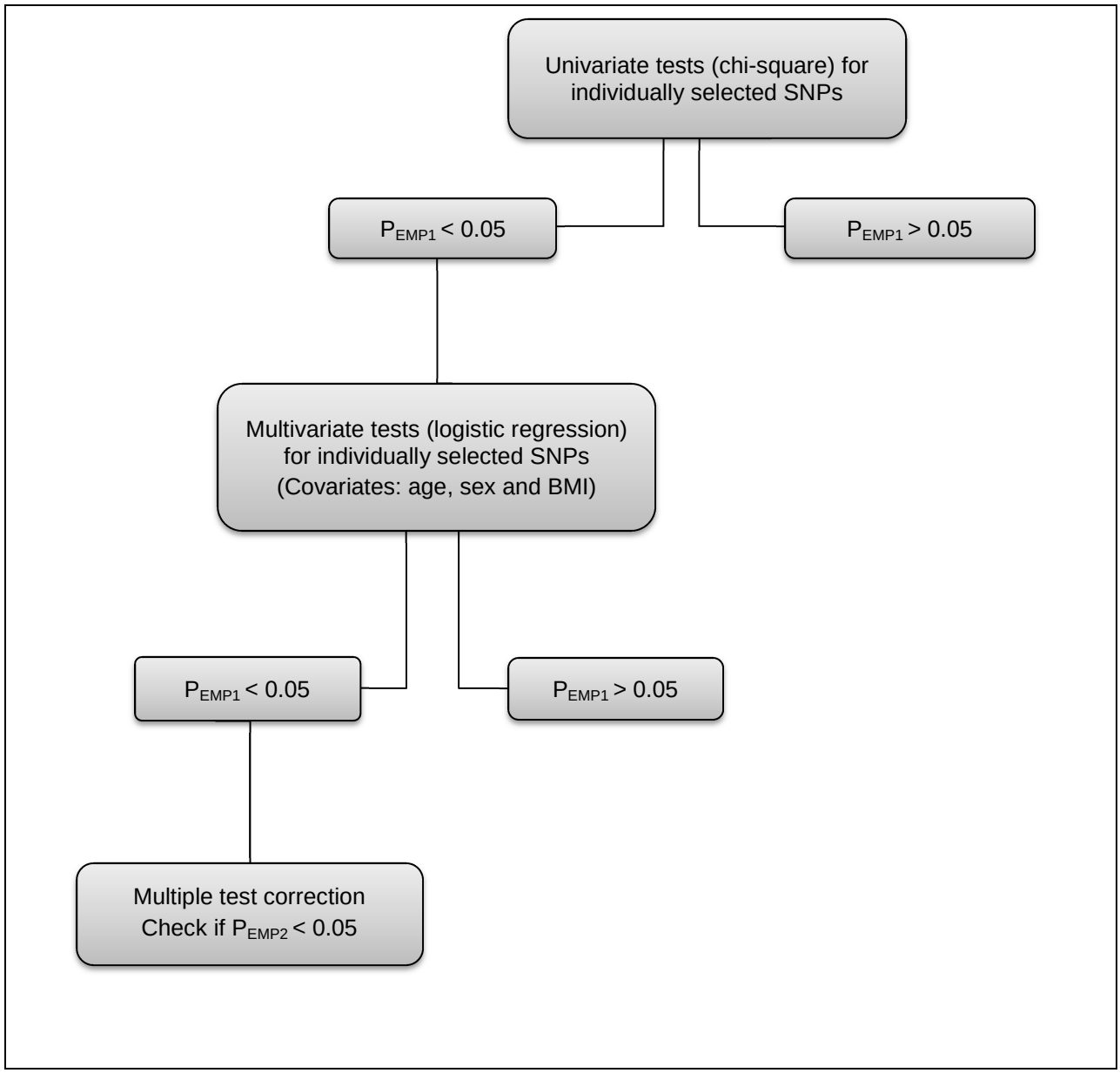
### 3.8.4 Analysis strategy for the genetic association tests

The genetic association tests were conducted using the analysis strategy illustrated in Figure 3.2 below. Initially, univariate tests were conducted using the chi square test for all the genotypic models of inheritance and the Fisher's exact test was used in cases where specific genotypes of the SNPs were carried by fewer than five people. Permutation tests were used to generate p-values as well as correcting for multiple testing based on the genotyped data set. The permutation tests are regarded as better suited for analysis of genetic association studies with small sample sizes (Clarke *et al*., 2011). Unlike the Bonferroni and Sidak corrections, which adjust for multiple testing

between markers by assuming independence tests, the permutation test is regarded as the gold standard for accurate correction of the p-values for multiple tests (Westfall & Young, 1993).

The SNPs that had significant differences indicated, as determined by a p value < 0.05, were further evaluated through the multivariate analysis. The logistic regression was used to evaluate association of the minor allele of each specific SNP with IGT while adjusting for age, sex and BMI, as illustrated in Figure 3.4. Using logistic regression, the odds ratio (OR) was calculated in reference to the minor allele in the study sample using the gPlink software (Purcell *et al*., 2007). An OR greater than 1 indicated that the SNP increased the risk of IGT while an OR less than 1 was regarded as indicating a protective effect against IGT. To correct for multiple testing bias, the family-wise correction for multiple comparisons ($P_{EMP2}$) value, based on 1000 permutations, was used and considered statistically significant if it was less than 0.05. Findings that had both $P_{EMP1}$ and $P_{EMP2}$ < 0.05 were considered statistically significant. SNPs with a minor allele frequency (MAF) of less than 0.01 were also removed as part of the quality control procedure to remove the possibility of misinterpreting false positive results of such SNPs from only a small number of participants in this study (Anderson *et al.,* 2010). It is recommended that rare variants are removed in studies of limited sample size, as a great deal of power is required to assess their association patterns (Morris *et al*., 2010).

**Figure 3.2 Analysis strategy for the genetic association tests**




$P_{EMP1}$ = point-wise empirical value; $P_{EMP2}$ = family-wise correction for multiple comparisons; SNPs = single nucleotide polymorphisms

# CHAPTER 4

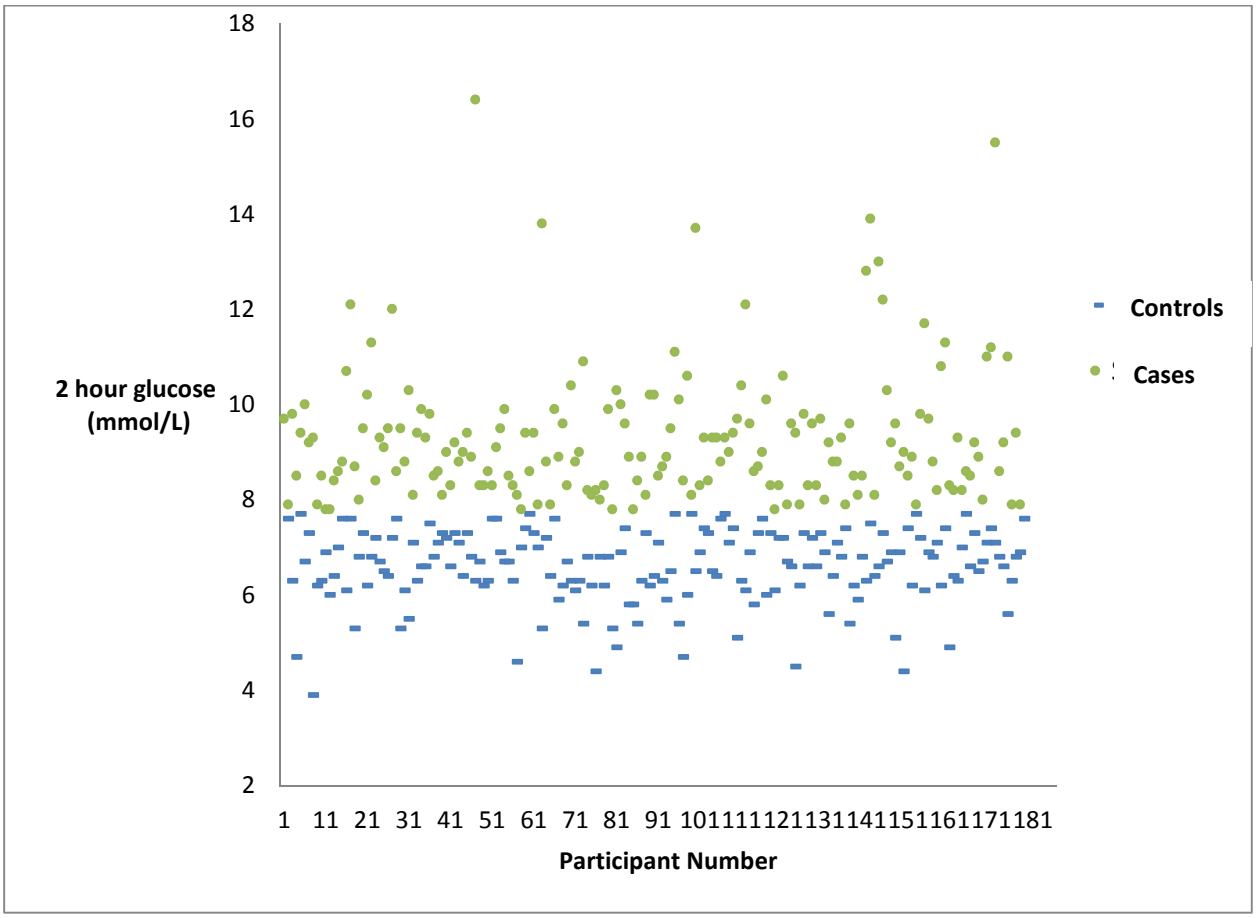
# RESULTS AND DISCUSSION

This study was undertaken to evaluate the association of specific common genetic variants that have previously been associated with T2D susceptibility in non-African populations with T2D risk, as indicated by IGT, in a black South African population. The following chapter discusses the results determined by using the methodology as outlined in Chapter 3. Section 4.1 will focus on the descriptive statistics of the research participants. Section 4.2 discusses the results of the DNA isolation and Section 4.3 provides a concise analysis of the BeadXpress® results clarifying how the quality of the results was assured before the genetic association tests were conducted. Section 4.4 explains the manner in which the HWE status of the selected SNPs was evaluated to ensure that the distribution of the SNPs was the result of random mating and not due to drift, selection, mutation, selective mating or migration. The genetic association tests are discussed in Section 4.5, together with how the study results compare with earlier studies. Finally Section 4.6 provides a summary of the significant findings of this study.

## 4.1 DESCRIPTIVE STATISTICS OF RESEARCH PARTICIPANTS

The descriptive statistics of the cases and controls included in the study sample were calculated and the results tabulated as displayed in Table 4.1. The descriptive statistics of the cases and controls were compared to identify confounding variables to adjust for in the genetic association analysis. The study participants had high carbohydrate intake and low fat intake as indicated in Table 4.1. However, there were no significant differences on the carbohydrate and fat intakes of the cases and controls even when the different types of fat such as monounsaturated fats, saturated fats and polyunsaturated fats were considered. However, the two-hour glucose concentrations, which were used to classify the cases and controls, BMI and urbanisation, were noted to be significantly different between these groups. A scatter plot of the two-hour glucose concentrations indicated that the the cases and controls had been correctly classified as a separation of the cluster was noted as indicated in Figure 4.1.

**Figure 4.1 Scatter plot for the 2 hour glucose concentrations of the study participants**




mmol/L = millimoles per litre

Although urbanisation was different among the cases and controls, it was not selected as a confounder. In contrast to classical epidemiological studies, factors such as urbanisation, dietary intake and other environmental factors are unlikely to confound the SNP association test, as they are unable to change genotypes themselves (Lunetta, 2008). However, population stratification as a result of people of different ethnicities residing in the same area has been reported to confound genetic association tests (Clarke *et al*., 2011). The research subjects were predominately black South Africans of Tswana descent, regardless of whether they resided in urban or rural areas, making the possibility of population stratification unlikely. Urbanisation was thus not regarded as a confounding variable for the genetic association tests in this study.

BMI was selected as a covariate to be adjusted for, as evidence of its confounding effect on the genetic associations of T2D exists. It has been reported that certain

SNPs, for example those in the FTO gene that are associated with BMI, tend to associate with T2D when BMI is not corrected for in genetic association analyses (Groop & Pociot, 2013). It is postulated that the FTO variants have an indirect association with T2D through BMI (Groop & Pociot, 2013); therefore BMI was selected as a covariate in order to correct for it and thus ensure that SNPs associated with T2D risk and not BMI were determined in this study.

The cases and controls were matched on an individual basis for age and sex to control for the confounding effects of these variables in the genetic association analysis. The paired t-test, McNemar's test and conditional logistic regression are regarded as the appropriate statistical approaches for matched case-control studies, as they factor in the matching during statistical analysis (Conway *et al*., 2013). However, the gPlink software, which was used to compute the genetic association tests, could not perform these analyses and therefore the independent t- test, chi square test and logistic regression tests were used, which do not factor in the effect of matching during statistical analysis (Conway *et al*., 2013). Statistical tests in matched case-control studies that do not factor in the effect of matching during analysis have been reported in a recent review by Conway *et al*., (2013) to result in inaccurate strength of association. Therefore, age and sex were adjusted for in the genetic association analysis to improve the accuracy of the statistical tests used in this study that were unable to factor in matching. The gPLINK software was used for the determination of genetic associations, as it is customised to evaluate the standard models of disease penetrance for SNPs, as opposed to other statistical software packages such as SPSS, as outlined in Section 3.8.

**Table 4.1 Descriptive statistics of the research participants**

| Variable | Cases | Controls | P-values |
|---|---|---|---|
| **Male (n)** | 39 | 39 | --- |
| **Female (n)** | 139 | 139 | --- |
| **Rural/Urban (n)** | 110/68 | 139/39 | 0.01 |
| **2 hr Glucose (mmol.L$^{-1}$)** | 9.30 ± 1.40 | 6.59 ± 0.78 | 0.00 |
| **Age (years)** | 42.88 ± 4.90 | 42.51 ± 5.16 | 0.48 |
| **Body mass index (kg/m$^2$)** | 25.92 ± 7.82 | 22.92 ± 6.02 | 0.00 |
| **Protein % of TE** | 12.17 ± 1.82 | 11.54 ± 1.88 | 0.92 |
| **Fat % of TE** | 23.56 ± 8.12 | 21.79 ± 7.54 | 0.30 |
| **Carbohydrate % of TE** | 63.98 ± 9.58 | 66.87 ± 9.26 | 0.60 |
| **MUFA % of TE** | 6.01 ± 3.13 | 5.39 ± 2.90 | 0.30 |
| **PUFA % of TE** | 7.28 ± 3 | 6.96 ± 3.07 | 0.86 |
| **SATFAT % of TE** | 5.62 ± 2.97 | 4.95 ± 2.60 | 0.86 |
| **Total fibre intake (g)** | 20.62 ± 9.45 | 20.76 ± 10.35 | 0.47 |
| **Total energy without alcohol (kJ)** | 7,428.17 ± 3,354.26 | 7,427.81 ± 3,338.02 | 0.65 |

Mean and standard deviations reported for the variables; TE = total energy; MUFA = mono-unsaturated fatty acids; PUFA = polyunsaturated fatty acids; SATFAT= saturated fatty acids; % = percentage; g = grams; kJ = kilojoules; kg = kilograms; m = metres; n = number; mmol = millimoles; L = litre

## 4.2 DNA ISOLATION

The DNA samples for this study were isolated using the FlexiGene™ DNA extraction kit (QIAGEN) as described in Section 3.7.3. The purity of the DNA was analysed using the NanoDrop® spectrophotometer as described in Section 3.7.4. The mean and standard deviation of the A260/A280 absorbance ratio of the DNA samples was 1.83 ± 0.07. The A260/A280 absorbance above 1.8 indicates that the DNA samples are of good quality. The mean yield of the genomic DNA after the extraction was noted to be 143.21ng.µL$^{-1}$. The DNA samples were then normalised to a final working dilution of concentration between 45-55 ng.µL$^{-1}$, which is required for genotyping with the Illumina® BeadXpress® platform.

## 4.3 ANALYSIS OF THE BEADXPRESS® RESULTS FOR SELECTED SNPS

Initially, a literature search of the reported common genetic variants associated with T2D in multiple ethnicities was undertaken, as elaborated on in Section 3.7.1. The selected SNPs were pre-assessed for their ability to be successfully genotyped with VeraCode® GoldenGate® technology from Illumina® using the ADT tool, as described in

Section 3.7.2. Genotyping kits were designed for the SNPs that had metrics that were indicative of high chances of successful genotyping. The SNPs were in turn genotyped using the Illumina® BeadXpress® Reader, as detailed in Section 3.7.5, after which the quality control procedures discussed in this section were conducted to ensure that the BeadXpress® results considered for the genetic association analysis were of suitable quality. The results of the procedures outlined above are discussed in the following sections.

### 4.3.1 Pre-assessment of SNPs for BeadXpress® genotyping

The Illumina® ADT tool generated a SNPScore file, which was used to assess the probability of successfully genotyping the selected SNPs for this study. The SNPScore file had metrics that were used to decide whether these SNPs could have genotyping assays designed for them using the VeraCode GoldenGate technology. The SNPScore file for the selected SNPs for this study is illustrated in Table 4.2. The SNP score, designability rank and validation status were the metrics used to evaluate the SNPs. The SNP score ranges from 0.00-1.00 and indicates the success of designing a viable assay for a specific SNP. The higher the SNP score, the greater the chances of successfully designing assay for the SNP. The designability rank is a metric that categorises the SNP score values. The ranks in this metric can be 0, 0.5 and 1. The SNPs with a low success rate for designing assay kits will have a rank of 0, which represents SNP score values lower than 0.4. SNPs with a moderate success rate will have a rank of 0.5, implying they have an SNP score between 0.4 and 0.6. SNPs with high SNP scores will have a value of 1, which indicates they have SNP scores between 0.6 and 1.1. The validation status was another metric that was evaluated from the SNPScore file. Three broad categories exist for this metric, namely GoldenGate validation status, two-hit validation status and non-validated. A SNP that was previously designed and successfully generated polymorphic results on the Illumina platform will be indicated as GoldenGate validated, signifying it has high chances of being genotyped successfully. A two-hit validation status implies that both alleles of the SNP have been determined either via two independent methods or in two different populations, implying it has a better probability of being genotyped well. Having non-validated status indicates that the SNP has a very low chance of being genotyped successfully and even if it might have a high designability score, there are chances that

the SNP might be monomorphic. Monomorphic SNPs have the same allele in the whole population and lack variation in that population. However, this is different for the polymorphic SNPs, which exhibit variation by having two different alleles in the population. Monomorphic SNPs can thus not be considered in genetic association tests owing to their lack of variation.

Upon evaluating the SNPScore file of the selected SNPs investigated in this study, it was noted that all the SNPs had designability ranks of either 0.5 or 1, implying that they had moderate to high success rates for successful assay design. However, when the metric for validation status was considered, rs12779790 and rs13266634 were noted to have non-validated status. These two SNPs were dropped in view of their metrics, which indicated a lower chance of genotyping them successfully, as discussed above. Therefore a final SNPScore file was compiled for the remaining 77 SNPs and sent to Illumina for the design of the assay kits.

**Table 4.2 Results of the pre- evaluation of the SNPs for BeadXpress® genotyping**

| Locus Name | Source | SNP Score | Designability Rank | Validation Bin |
|---|---|---|---|---|
| rs10460009 | dbSNP | 0.916 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs1048886 | dbSNP | 1.1 | 1 | GoldenGate Validated |
| rs10741243 | dbSNP | 0.806 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs10923931 | dbSNP | 0.883 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs1111875 | dbSNP | 1.1 | 1 | GoldenGate Validated |
| rs1153188 | dbSNP | 0.774 | 1 | TwoHit_OneKGenomeValidated |
| rs11634397 | dbSNP | 0.971 | 1 | HapMap_OneKGenomeValidated |
| rs12027542 | dbSNP | 1.1 | 1 | GoldenGate Validated |
| rs12518099 | dbSNP | 0.699 | 1 | HapMap_OneKGenomeValidated |
| rs12779790 | dbSNP | 0.423 | 0.5 | Non-validated |
| rs13266634 | dbSNP | 0.871 | 1 | Non-validated |
| rs13292136 | dbSNP | 0.892 | 1 | OneKGenome Validated |
| rs1359790 | dbSNP | 0.967 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs1387153 | dbSNP | 0.509 | 0.5 | TwoHit_OneKGenomeValidated |
| rs1436955 | dbSNP | 0.884 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs1531343 | dbSNP | 0.977 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs1552224 | dbSNP | 1.1 | 1 | GoldenGate Validated |

| Locus Name | Source | SNP Score | Designability Rank | Validation Bin |
|---|---|---|---|---|
| rs16861329 | dbSNP | 0.94 | 1 | OneKGenome Validated |
| rs16955379 | dbSNP | 0.822 | 1 | OneKGenome Validated |
| rs17036101 | dbSNP | 0.891 | 1 | OneKGenome Validated |
| rs17045328 | dbSNP | 0.829 | 1 | OneKGenome Validated |
| rs17584499 | dbSNP | 0.91 | 1 | OneKGenome Validated |
| rs17797882 | dbSNP | 0.942 | 1 | OneKGenome Validated |
| rs1801214 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs1801282 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs1802295 | dbSNP | 0.947 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs2028299 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs2063640 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs2237892 | dbSNP | 0.791 | 1 | OneKGenomeValidated |
| rs2237897 | dbSNP | 0.499 | 0.5 | OneKGenomeValidated |
| rs231362 | dbSNP | 0.621 | 1 | TwoHit_OneKGenomeValidated |
| rs243021 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs2722769 | dbSNP | 0.894 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs2833610 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs2943641 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs358806 | dbSNP | 0.804 | 1 | HapMap_OneKGenomeValidated |
| rs3773506 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs3786897 | dbSNP | 0.792 | 1 | TwoHit_OneKGenomeValidated |
| rs3813627 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs391300 | dbSNP | 0.46 | 0.5 | TwoHit_OneKGenomeValidated |
| rs3923113 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs4402960 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs4430796 | dbSNP | 0.956 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs4457053 | dbSNP | 0.836 | 1 | TwoHit_OneKGenomeValidated |
| rs4607103 | dbSNP | 0.909 | 1 | TwoHit_OneKGenomeValidated |
| rs4689388 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs472265 | dbSNP | 0.697 | 1 | TwoHit_OneKGenomeValidated |
| rs4812829 | dbSNP | 0.985 | 1 | OneKGenomeValidated |
| rs5082 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs564398 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs5945326 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs6467136 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs6583826 | dbSNP | 1.1 | 1 | GoldenGateValidated |

| Locus Name | Source | SNP Score | Designability Rank | Validation Bin |
|---|---|---|---|---|
| rs6815464 | dbSNP | 0.811 | 1 | OneKGenomeValidated |
| rs7018475 | dbSNP | 0.877 | 1 | TwoHit_OneKGenomeValidated |
| rs7041847 | dbSNP | 0.791 | 1 | TwoHit_OneKGenomeValidated |
| rs7107217 | dbSNP | 0.843 | 1 | TwoHit_OneKGenomeValidated |
| rs7172432 | dbSNP | 0.979 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs7178572 | dbSNP | 0.914 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs730570 | dbSNP | 0.98 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs7542900 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs7560163 | dbSNP | 0.884 | 1 | TwoHit_OneKGenomeValidated |
| rs7578326 | dbSNP | 0.885 | 1 | TwoHit_OneKGenomeValidated |
| rs7578597 | dbSNP | 0.906 | 1 | OneKGenomeValidated |
| rs7593730 | dbSNP | 0.693 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs7630877 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs7636 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs7754840 | dbSNP | 0.921 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs7903146 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs7957197 | dbSNP | 0.564 | 0.5 | OneKGenomeValidated |
| rs7961581 | dbSNP | 0.965 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs8042680 | dbSNP | 0.826 | 1 | TwoHit_OneKGenomeValidated |
| rs8050136 | dbSNP | 0.941 | 1 | TwoHit_HapMap_OneKGenomeValidated |
| rs831571 | dbSNP | 0.983 | 1 | HapMap_OneKGenomeValidated |
| rs864745 | dbSNP | 0.875 | 1 | TwoHit_OneKGenomeValidated |
| rs896854 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs9470794 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs9472138 | dbSNP | 1.1 | 1 | GoldenGateValidated |
| rs972283 | dbSNP | 0.808 | 1 | TwoHit_OneKGenomeValidated |

### 4.3.2 Analysing data from the BeadXpress® reader

The DNA samples were genotyped for the selected common genetic variants as described in Section 3.7.5, after which a quality control procedure was undertaken to ensure that the BeadXpress® results were of suitable quality. Initially, the controls of the BeadXpress® genotyping process were evaluated to check if the reaction had occurred as expected. After this the clustering of the outputs generated from the genotyping reaction was evaluated to ensure that the selected SNPs had been genotyped successfully. The BeadXpress® genotyping was validated by randomly selecting

samples for particular SNPs and sequencing them as described in Section 3.7.6. The samples which had genotypes within the cluster and those with genotypes between the clusters were selected to validate the genotype calling of the BeadXpress reader. The validation was prioritised for the few samples which had unclear clusters, thus few samples were validated. The quality control procedure that was performed to analyse the BeadXpress® results is discussed in detail in the following sections.

### 4.3.3 Evaluation of the BeadXpress® Reader genotyping controls

The BeadXpress® genotyping process contains a number of internal controls that help to give the assurance that the reaction occurs successfully and that can be used for troubleshooting of the process. The controls that were evaluated are allele-specific extension controls, contamination controls, PCR uniformity tests, extension gap controls as well as first hybridisation and second hybridisation controls. These controls are recognised by their IllumiCode Sequence IDs and are expected to yield fluorescence in either the Cy3 or Cy5 fluorescent channel of the BeadXpress® apparatus to indicate that the genotyping process was successful. The fluorescence yielded by the controls should match the expected fluorescence of that particular control. When a control yields the expected fluorescence in the Cy3 channel, it is termed a U3 match and when a control yields the expected fluorescence in the Cy5 channel, it is termed a U5 match. The expected U3 and U5 matches of the controls are indicated in Table 4.3 below. The detailed results of the genotyping controls are discussed in subsequent sections.

**Table 4.3 IllumiCode sequence IDs of controls and their expected outcomes**

| Class of controls | IllumiCode sequence ID | Description | Expected outcome |
|---|---|---|---|
| Allele-specific extension controls | 329 | AA mismatch | U3 match |
| | 1611 | CC mismatch | U5 match |
| | 1142 | GG mismatch | U3 match |
| | 279 | GT mismatch | U5 match |
| Polymerase chain reaction controls | 1742 | High AT (31% GC) | U3 match |
| | 4824 | High GC (62% GC) | U5 match |
| Extension gap controls | 658 | 15-bp gap | U3 and U5 match |
| First hybridisation controls | 962 | First hybridisation controls, 42/57 Tm | U5 match |
| | 1209 | First hybridisation controls, 57/72 Tm | U5 match |
| Second hybridisation | 44 | Second hybridisation controls | U3 match |

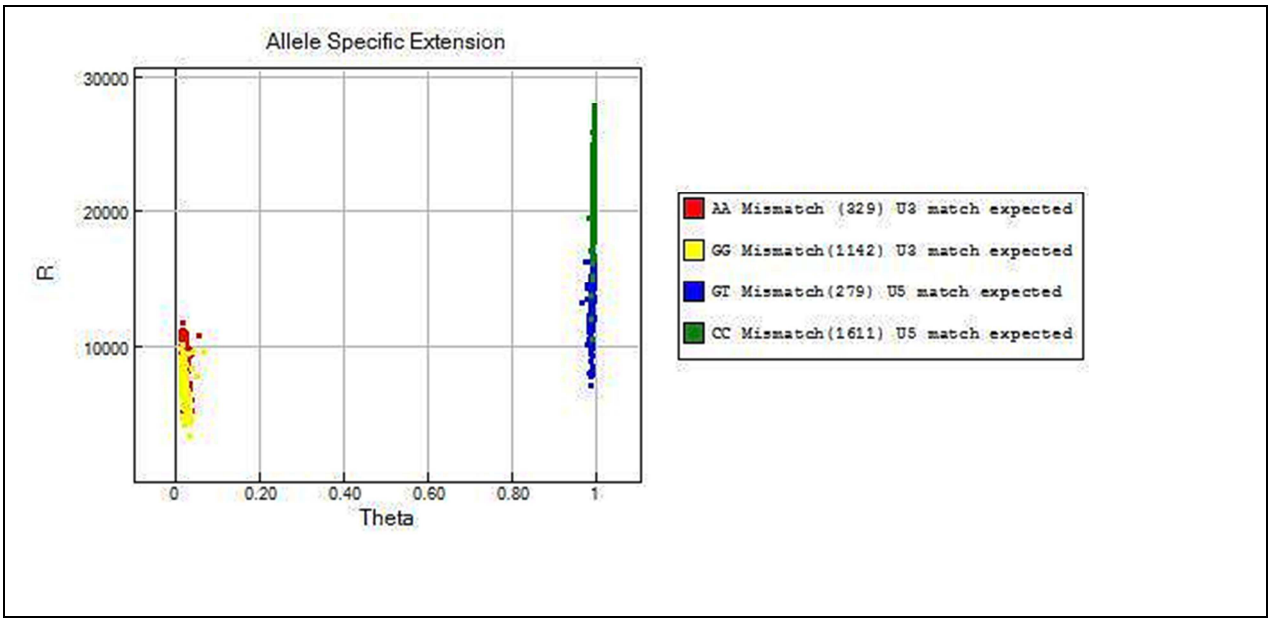
| | | | |
|---|---|---|---|
| controls | 278 | Second hybridisation controls | U3 match |
| | 1112 | Second hybridisation controls | U5 match |
| | 1632 | Second hybridisation controls | U5 match |
| | 501 | Second hybridisation controls | U3 and U5 match |
| | 1003 | Second hybridisation controls | U3 and U5 match |

a = adenine; bp = base pairs; c = cytosine; g = guanine; t = thymine; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel; $T_m$ = melting temperature; $T_a$ = annealing temperature

#### 4.3.3.1 Allele-specific extension controls

During the BeadXpress® genotyping reaction there are two ASOs that are designed specifically to hybridise to either of the two alleles of the selected SNPs being investigated. The allele-specific extension controls measure the efficiency of the hybridisation of the properly matched ASO. Mismatches for A-A, C-C, G-N-BG, and G-T corresponding to IllumiCode sequence IDs 329, 1611, 1142 and 279 respectively are measured. The sequence IDs 329 and 1142 are expected to be predominantly Cy3 and sequence IDs 279 and 1611 should be predominantly Cy5. As indicated in Figure 4.2, all the sequence IDs had the expected U3 and U5 matches, thus indicating that the ASOs were able to hybridise correctly to the specific alleles of SNPs that were genotyped.

**Figure 4.2 Allele-specific extension controls**

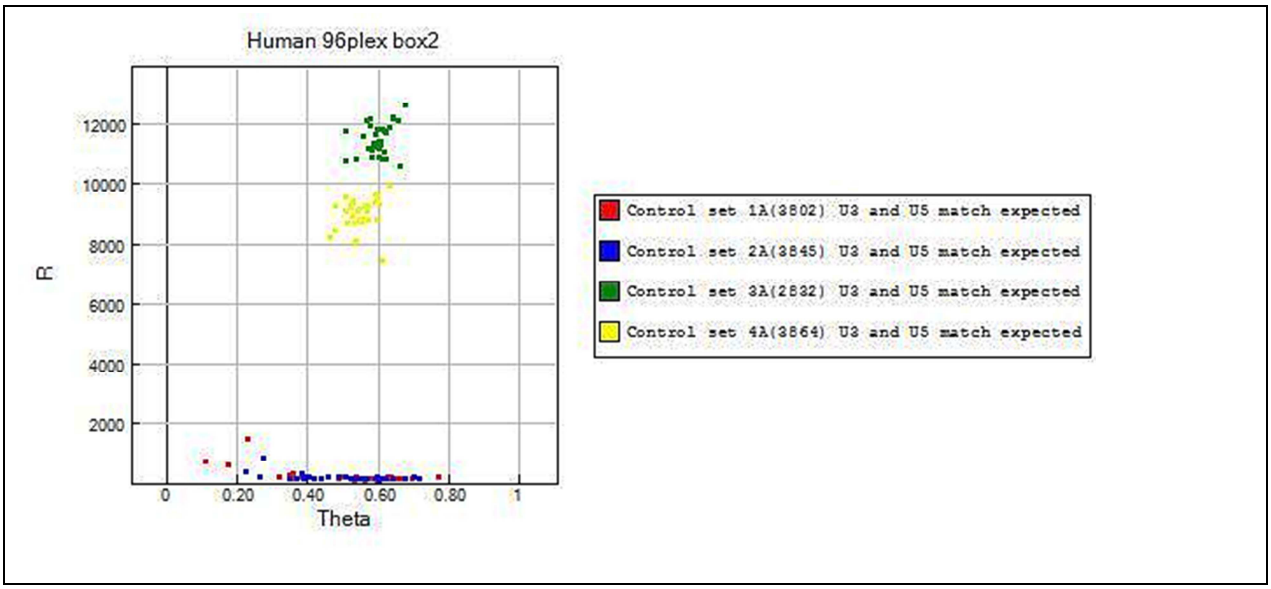



A = adenine; C = cytosine; G = guanine; T = thymine; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

### 4.3.3.2 Contamination controls

Contamination controls are also built into the BeadXpress® genotyping assay to ensure that no carry-over contamination occurs in the plates that contain the DNA samples. Figure 4.3 is a representation of the contamination controls for a single plate. Lack of contamination is indicated by the U3 and U5 matches. It is also representative of the graphs for all the plates and therefore indicates that no plate had any carry-over contamination, since the expected U3 and U5 matches were noted as illustrated in Figure 4.3.

**Figure 4.3 Contamination controls**

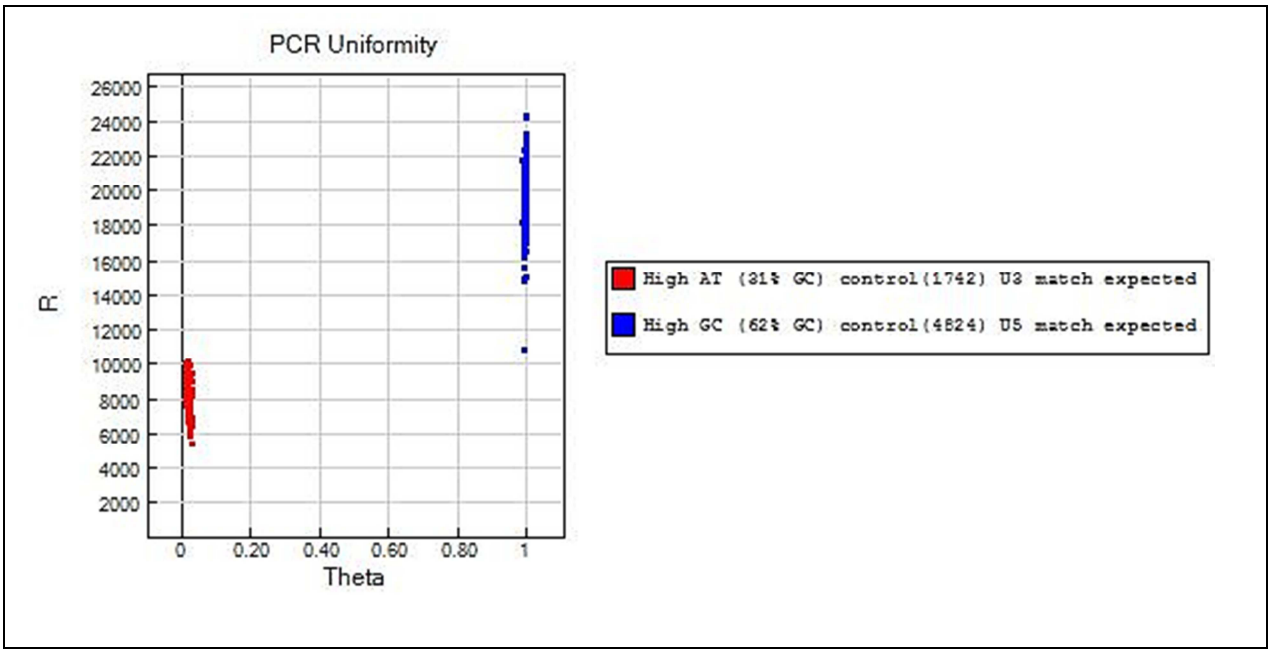



U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

#### 4.3.3.3 Polymerase chain reaction uniformity controls

The PCR uniformity controls test the efficiency of the amplification process for high AT-rich and high GC-rich regions of DNA. The amplification of the high AT-rich (31% GC) DNA regions is assessed by evaluating the IllumiCode Sequence ID 1742 and for the high GC-rich regions the IllumiCode Sequence ID 4824 is used. The IllumiCode Sequence ID 1742 and 4824 had U3 and U5 matches respectively, as expected. Thus the expected PCR amplifications for the high AT and GC regions of the DNA were attained during the BeadXpress® genotyping process and this also indicated that the other regions, which are less difficult to amplify compared to the high AT and GC regions, were amplified successfully. Therefore, through the results of these controls, assurance was provided that the amplification process had occurred successfully as indicated in Figure 4.4.

**Figure 4.4 Polymerase chain reaction uniformity controls**

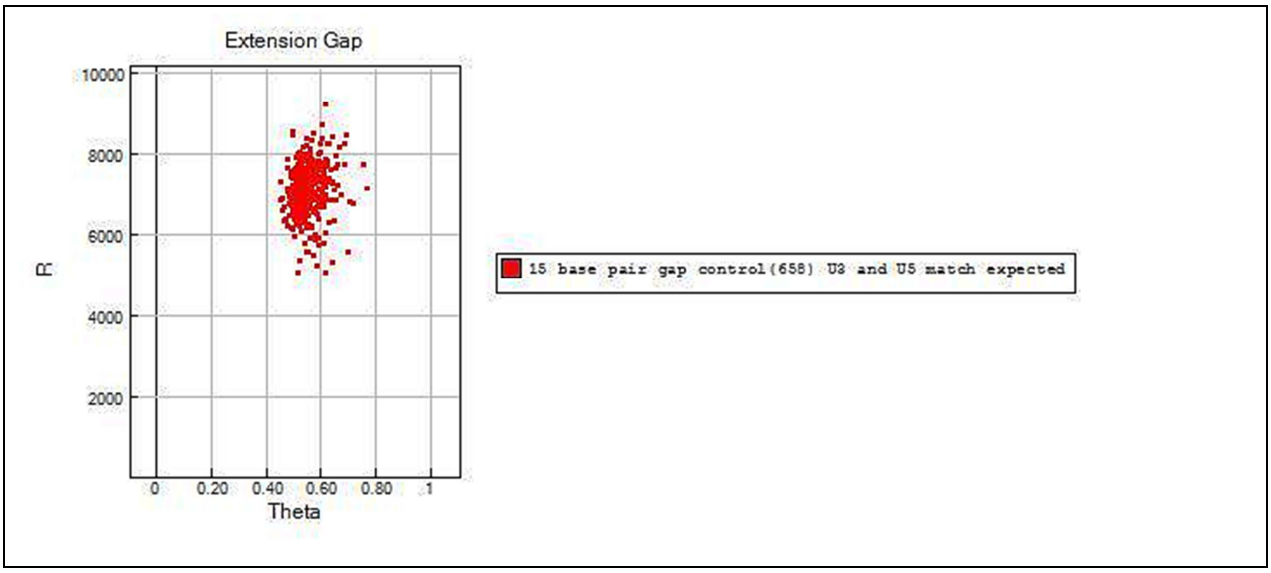



A = adenine; C = cytosine; T = thymine; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

#### 4.3.3.4 Extension gap controls

The extension gap control evaluates the efficiency of the extension of the 15 bases from the 3' end of the ASO to the 5' end of the locus-specific oligonucleotide. For this control both Cy3 and Cy5 signals are expected to be detected through their respective channels. Figure 4.5 indicates that both the Cy3 and Cy5 signals were detected as expected. Therefore the extension phase during the amplification of the selected SNPs was successful, as evidenced by the expected results of the extension gap controls.

**Figure 4.5 Extension gap controls**

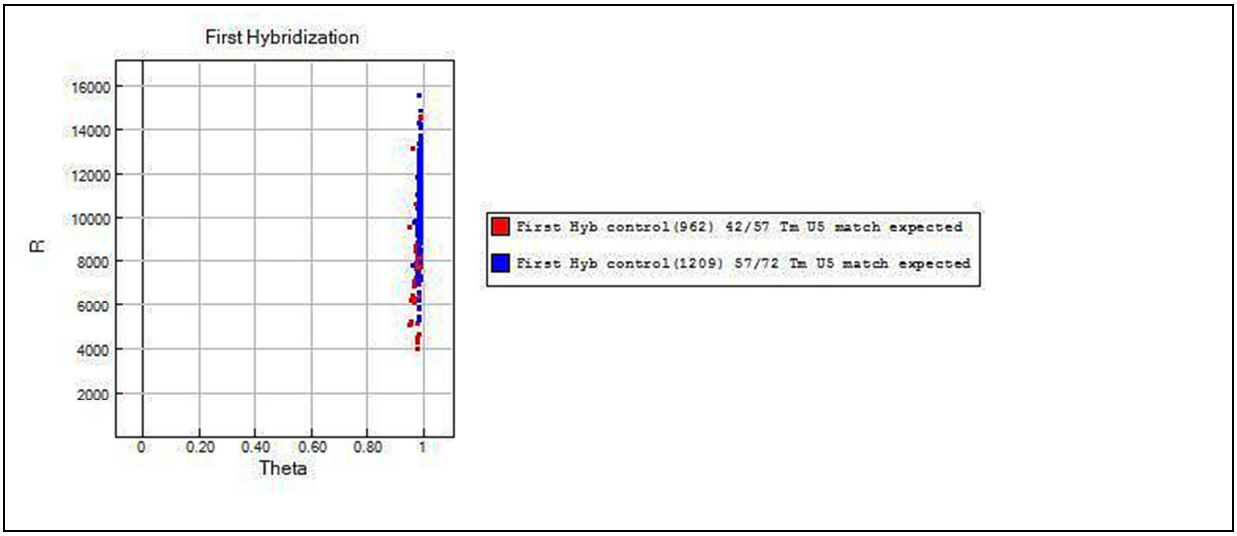



bp = base pair; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

#### 4.3.3.5 First hybridisation controls

The specificity of the annealing of the ASOs at a different melting temperature ($T_m$) to the same DNA locus is evaluated by the first hybridisation controls. The IllumiCode Sequence ID 962 and 1209 were evaluated for this reason and were expected to yield Cy5 signals, which were depicted as U5 matches. Figure 4.6 indicates that the expected specificity of the ASO was attained, thus giving the assurance that the ASOs hybridised with specificity at the different melting temperatures.

**Figure 4.6 First hybridisation controls**

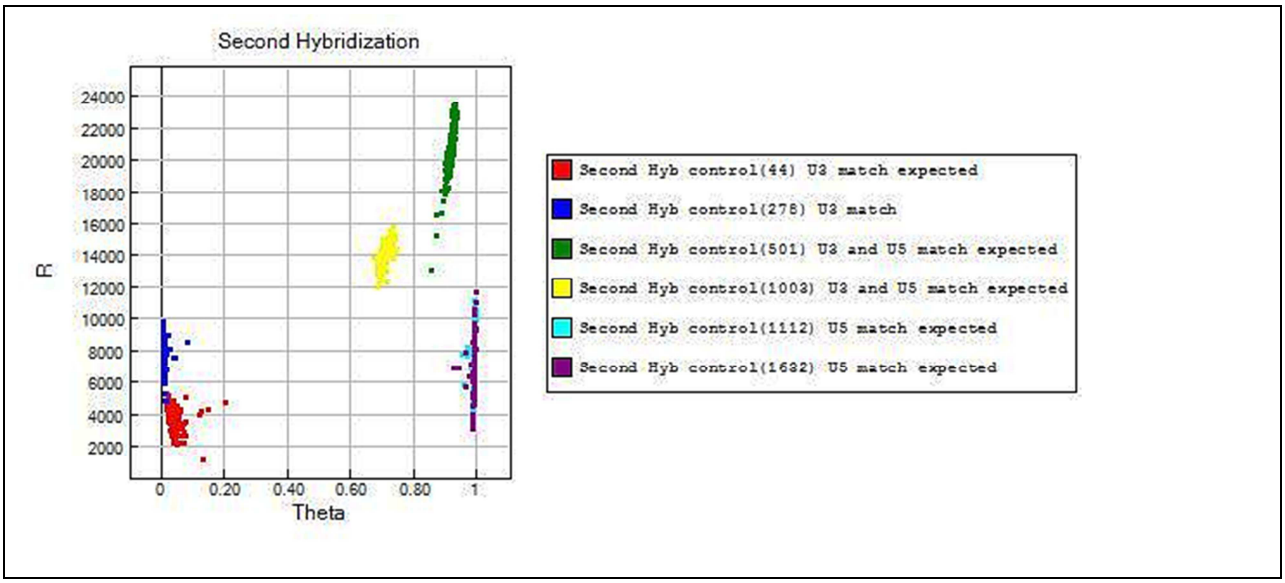



$T_m$ = melting temperature; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

#### 4.3.3.6 Second hybridisation controls

The second hybridisation controls assess the hybridisation of the single-stranded assay products to IllumiCode Sequences on the array beads. The Cy3 signal is expected for IllumiCode Sequence IDs 44 and 278, while the Cy5 signal should be emitted only for the 1112 and 1632 Sequence IDs. A signal that consists of the Cy3 and Cy5 is expected for the Sequence IDs 501 and 1003. Figure 4.7 indicates that all the expected signals were noted for all the second hybridisation controls and this gives the assurance that the amplification process went well.

**Figure 4.7 Second hybridisation controls**




Hyb = hybridisation; U5 = refers to the Cy5 channel; U3 = refers to the Cy3 channel

### 4.3.4 Evaluation of the cluster plots of the selected SNPs

A quality control procedure was conducted to ensure that the genotypes of the selected SNPs were well clustered by evaluating a number of metrics. The GenCall Score, call rate, mean normalised intensity (AB R Mean) and normalised theta (AB T Mean) values were the metrics used to assess the genotype clusters of the selected SNPs, as discussed below.

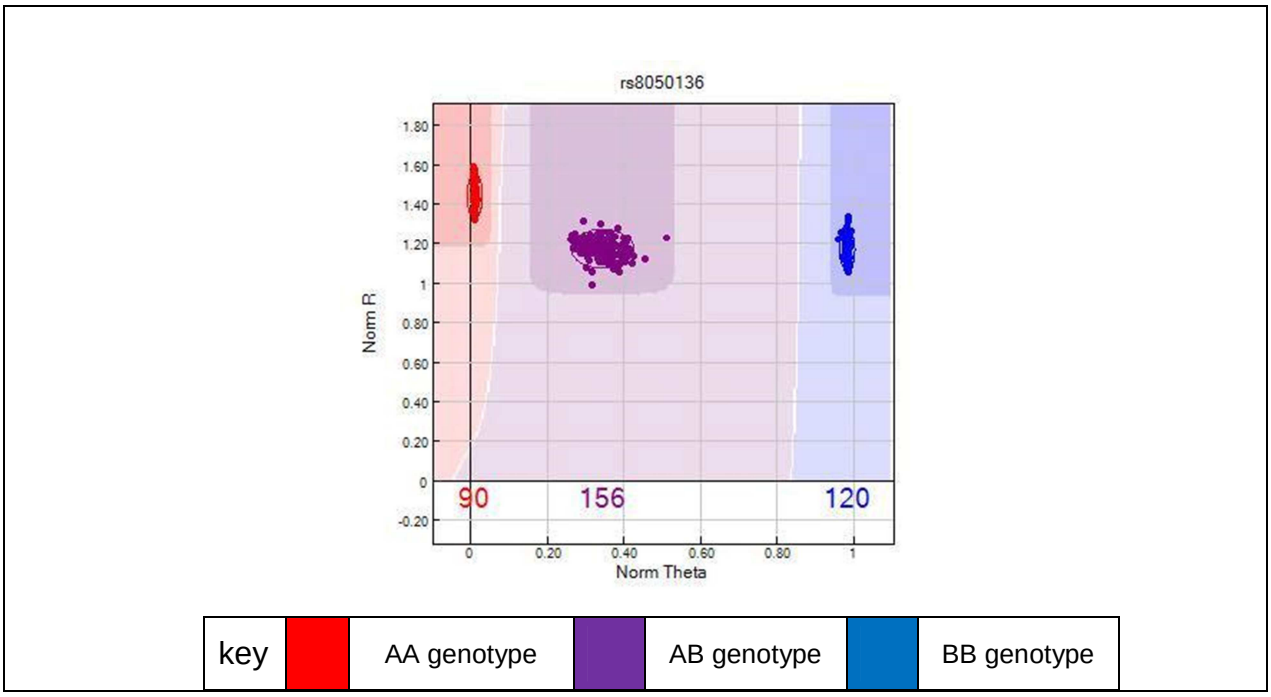
The GenCall score ranges from 0 to 1 and it is a measure of the distance a specific genotype call is from the centre of the cluster. The GenCall score is based on the

angle, dispersion, overlap and intensity of the SNP cluster and these are the characteristics used to determine this score by the GenomeStudio program. Samples with low GenCall scores are situated further from the centre of the cluster and hence are less reliable. In this study five samples that had GenCall scores below 0.5 were thus removed.

The call rate indicates the proportion of samples with scores above the no-call threshold for a specific SNP. The call rate values range from 0 to 1 and during the quality control, SNPs with call rates below 0.9 were not considered for the genetic association tests. The call rate cut-off was determined after reviewing the genotype clusters of the selected SNPs. Eleven SNPs that had call rates below 0.9 were thus removed from the genetic association tests, leaving 66 SNPs.

The AB R and the AB T of the GenomeStudio clustered plots were evaluated. The AB R mean and the AB T mean are metrics that indicate how well the genotypes of the heterozygote group are clustered. AB R mean values below 0.2 and AB T mean values lower than 0.2 and those in the range of 0.8-1 are indicative of poor clustering. The cluster plot of rs8050136 in Figure 4.8 was among those that were approved for genetic association analysis, as it had an AB R mean of 1.17 and an AB T mean of 0.34 that was indicative of good clustering.

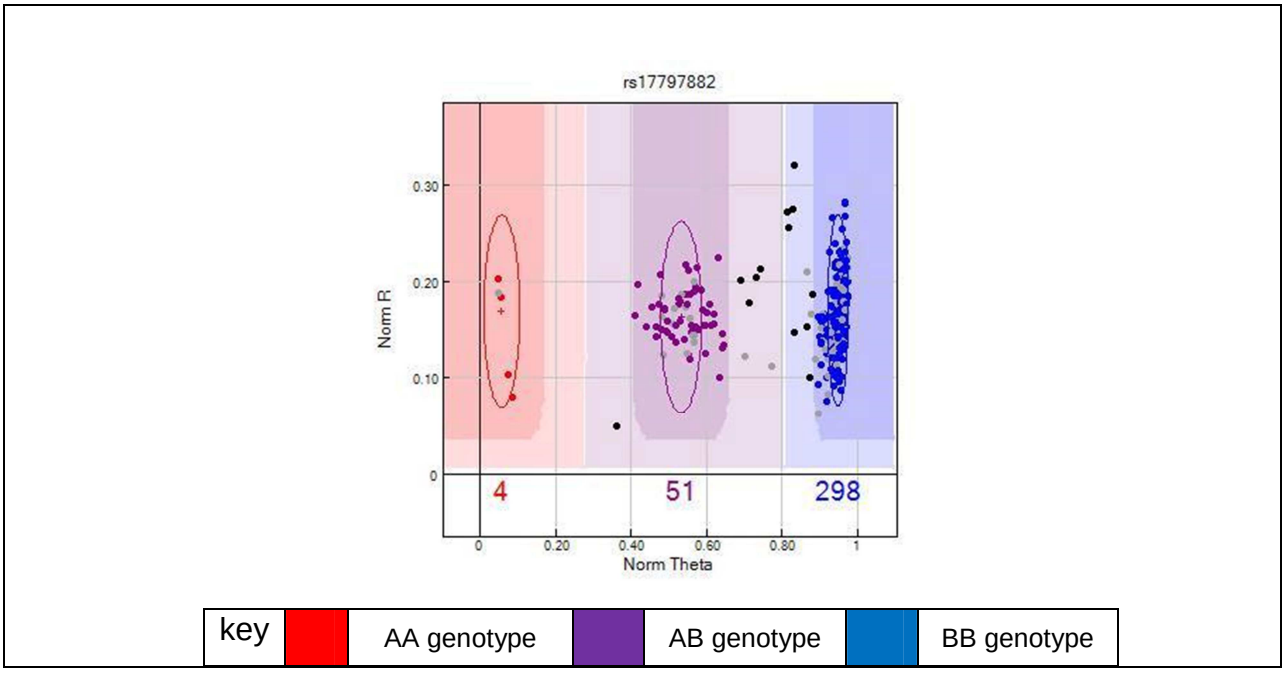
**Figure 4.8 Genotype cluster plot for the SNP rs8050136**




AA =homozygous mutant; AB = heterozygous; BB =homozygous, Norm R = norm intensity

The cluster plot of SNP rs17797882 indicated in Figure 4.9 was among those considered to have poor clustering, as it had an AB R mean lower than 0.2 and was thus excluded from the genetic association tests.

**Figure 4.9 Genotype cluster plot for the SNP rs17797882**




AA = homozygous mutant; AB = heterozygous; BB = homozygous, Norm R = norm intensity

### 4.3.5 Validation of the BeadXpress® genotyping

Random samples were selected for sequencing to validate the BeadXpress® results using the Sanger sequencing method as described in Section 3.8.3. Initially the PCR amplification reactions were optimised for the selected SNPs. During the optimisation stages the optimum temperature for the PCR amplification for the selected SNPs was determined and the expected product size evaluated. The PCR amplification of the randomly selected samples was performed using the conditions indicated in Section 3.8.2 and the primers illustrated in Section 3.8.1. After the amplification, the PCR products were sent for sequencing to the Central Analytical Facility in Stellenbosch. The sequences were aligned and the genotypes of these samples compared with those indicated by the GenomeStudio. There was 100% concordance between the sequencing results and the BeadXpress® results, as indicated in Table 4.4.

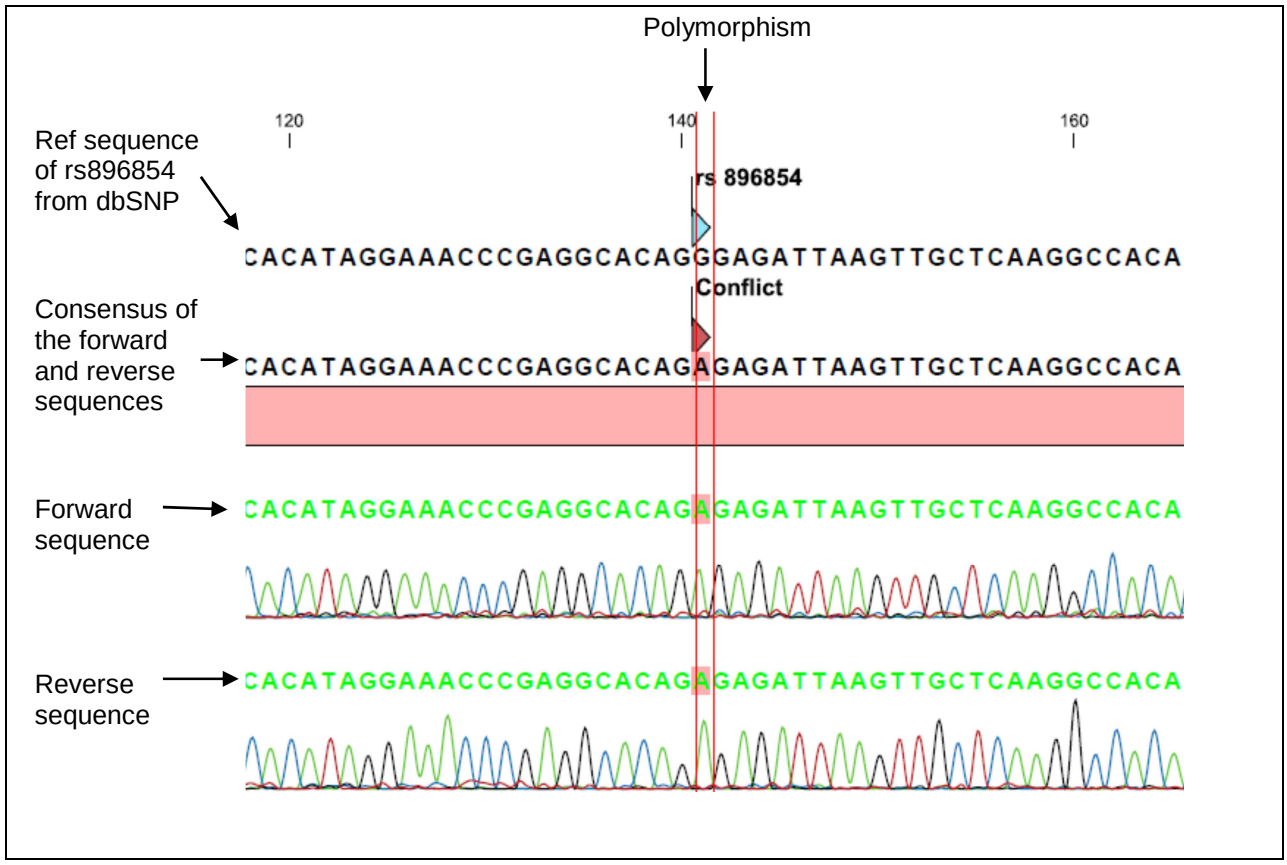
**Table 4.4 Comparisons of the BeadXpress genotyping and Sanger sequencing results for selected samples**

| Sample ID | SNP | Genotype according to BeadXpress | Genotype according to Sanger sequencing |
|---|---|---|---|
| 3076 | rs6467136 | GG | GG |
| 3042 | rs896854 | AA | AA |
| 3091 | rs896854 | AA | AA |
| 3022 | rs7630877 | GG | GG |

A = adenine; C = cytosine; G = guanine; rs = reference sequence; T = thymine

Figure 4.10 is an example of the electropherogram of the homozygous mutant genotypes that was validated by the Sanger sequencing method, which is described in detail in Section 3.6.6.

**Figure 4.10 Sequence electropherogram for the SNP rs896854 in individual 3091**




A = adenine; C = cytosine; G = guanine; rs = reference sequence; T = thymine

## 4.4 HARDY-WEINBERG EQUILIBRIUM STATUS OF THE SELECTED COMMON VARIANTS

After conducting the quality control procedures as previously discussed, the HWE status of the selected variants was evaluated. As described in Section 3.9.2, the HWE status was evaluated to ensure that the distribution of the genotypes of the selected SNPs was not a result of other factors such as genetic drift, migration and natural selection, but a result of random mating, which are the assumptions of the HWE model. The deviation from HWE is also indicative of genotyping error (Hosking *et al*., 2004).

The HWE status for the control groups is usually considered and not that of the cases in case-control studies, to avoid removing SNPs that are selected for by the disease among the cases (Anderson *et al*., 2010). Thus for this study, the HWE status of the controls was considered for all the models of genetic association tests except for the additive genetic model, for which the HWE status for both the cases and controls was

considered. The genetic models of association, such as the additive model work, on the assumption that the SNP of interest should be in HWE in the cases and also in the controls, hence before these tests were conducted it was essential to evaluate the HWE status of the selected SNPs in the cases and controls (Anderson *et al.,* 2010*)*.

The HWE status of the selected SNPs was evaluated by comparing the observed genotypes with the expected genotypes using the chi square method, as detailed in Section 3.8.2. Most of the SNPs adhered to the assumptions of HWE as indicated by p-values greater than 0.05 among both the cases and controls, as depicted in Table 4.2. However, three SNPs, i.e. rs7754840 ($p < 0.001$), rs13266634 ($p < 0.001$) and rs7542900 ($p = 0.03$), deviated from the HWE assumptions among the controls. Among the cases, the genotypic distributions of the following SNPs were noted to deviate from the assumptions of HWE: rs8050136 ($p < 0.001$), rs7754840 ($p < 0.001$), rs7961581 ($p = 0.02$) and rs7560163 ($p = 0.02$). In genetic association analyses, SNPs that do not adhere to the assumptions of the HWE are usually removed, as possible genotyping error is assumed to have occurred. However, according to Clarke *et al*. (2010), the SNPs that deviate from HWE can be retained if they have acceptable SNP genotyping clusters (Clarke *et al*., 2010). In view of the fact that the quality control procedures had been performed as discussed and 11 SNPs that had poor genotyping metrics had been removed as described in Section 4.3.4, the SNPs that had modest HWE deviations ($p = 0.02 – 0.03$), rs7542900, rs7961581 and rs7560163, were retained for genetic association tests on the basis of being well genotyped. However, the genotype frequencies of the SNPs that consisted of rs8050136, rs7754840 and rs1366634 that had large deviations of HWE ($p<0.001$) were evaluated to ascertain the probable reason for the deviance.

Upon evaluating the genotype frequencies of rs8050136 indicated in Table 4.4, it was noted that among the cases there was a reduction in the number of heterozygotes, while the minor allele frequencies increased. This was suggestive of the minor allele being selected for owing to the T2D status of the participants, thus being the probable reason for its departure from HWE among the cases. However, the deviation of rs13266634 from HWE among the controls was noted to be most likely because of genetic drift, as the genotypic frequencies indicated an increase in the major allele

accompanied by a decrease in the number of heterozygotes and the minor allele, as indicated in Table 4.4. However these reasons will need to be validated in a large sample size. The genotype frequencies of the SNP rs7754840 indicated an increase in the number of heterozygotes both among the cases and controls. Therefore, probable reasons for the deviations from HWE for this SNP might have been the poor genotyping of a highly homologous region for this SNP, which might have led to an increase in the low-copy repeats that have been noted to increase the proportions of heterozygotes (Wigginton *et al.,* 2005). This SNP was then removed from the genetic association tests, since it might have been poorly genotyped.

**Table 4.5 HWE status of the selected common variants**

| SNP ID | Major allele | Minor allele | MAF | Genotype Freq Cases XX/XY/YY | Genotype Freq in Controls XX/XY/YY | Cases HWE (P- Value) | Controls HWE (P- Value) |
|---|---|---|---|---|---|---|---|
| rs10460009 | T | C | 0.06 | 1/20/142 | 1/17/152 | 0.53 | 0.41 |
| rs1048886 | G | A | 0.30 | 11/75/77 | 14/72/84 | 0.26 | 0.86 |
| rs10741243 | G | C | 0.30 | 11/73/77 | 15/74/79 | 0.34 | 0.86 |
| rs10923931 | T | G | 0.24 | 10/50/103 | 11/65/94 | 0.25 | 1.00 |
| rs1111875 | T | C | 0.23 | 11/50/102 | 13/53/104 | 0.17 | 0.13 |
| rs1153188 | A | T | 0.19 | 8/41/112 | 5/58/100 | 0.11 | 0.48 |
| rs11634397 | G | A | 0.41 | 34/71/58 | 24/89/57 | 0.20 | 0.34 |
| rs12518099 | G | A | 0.29 | 7/72/83 | 12/78/76 | 0.11 | 0.21 |
| rs12779790 | G | A | 0.11 | 2/29/131 | 2/39/128 | 0.67 | 1.00 |
| rs13266634 | T | C | 0.10 | 4/24/126 | 9/15/138 | 0.06 | <0.001 |
| rs13292136 | T | C | 0.05 | 2/12/149 | 0/14/156 | 0.05 | 1.00 |
| rs1359790 | A | G | 0.05 | 1/15/145 | 0/15/155 | 0.36 | 1.00 |
| rs1387153 | T | C | 0.24 | 7/55/101 | 11/70/89 | 1.00 | 0.70 |
| rs1436955 | C | T | 0.42 | 35/80/47 | 22/87/60 | 1.00 | 0.33 |
| rs1531343 | G | C | 0.41 | 36/67/57 | 31/67/68 | 0.08 | 0.05 |
| rs16861329 | T | C | 0.02 | 0/6/157 | 0/8/162 | 1.00 | 1.00 |
| rs16955379 | T | C | 0.18 | 3/55/105 | 6/48/116 | 0.21 | 0.79 |
| rs17036101 | A | G | 0.13 | 1/36/126 | 4/43/123 | 0.70 | 1.00 |
| rs17584499 | T | C | 0.07 | 1/22/140 | 0/24/138 | 0.60 | 1.00 |
| rs17797882 | T | C | 0.08 | 0/19/136 | 3/28/134 | 1.00 | 0.38 |
| rs2028299 | C | A | 0.30 | 15/74/74 | 14/66/90 | 0.72 | 0.70 |
| rs231362 | A | G | 0.16 | 5/47/109 | 5/42/123 | 1.00 | 0.55 |
| rs243021 | A | G | 0.46 | 27/94/42 | 34/87/49 | 0.06 | 0.76 |
| rs2722769 | C | G | 0.02 | 0/7/156 | 0/4/166 | 1.00 | 1.00 |
| rs2833610 | A | G | 0.35 | 17/82/64 | 21/72/76 | 0.24 | 0.61 |

| SNP ID | Major allele | Minor allele | MAF | Genotype Freq Cases XX/XY/YY | Genotype Freq in Controls XX/XY/YY | Cases HWE (P- Value) | Controls HWE (P- Value) |
|---|---|---|---|---|---|---|---|
| rs2943641 | T | C | 0.34 | 19/66/64 | 14/72/65 | 0.72 | 0.46 |
| rs358806 | A | C | 0.04 | 0/7/119 | 0/13/104 | 1.00 | 1.00 |
| rs3773506 | G | C | 0.12 | 3/37/123 | 3/34/133 | 1.00 | 0.71 |
| rs391300 | C | T | 0.32 | 19/76/68 | 12/77/81 | 0.86 | 0.36 |
| rs3923113 | A | C | 0.15 | 5/39/119 | 3/45/122 | 0.37 | 0.77 |
| rs4402960 | G | T | 0.44 | 31/84/48 | 30/85/55 | 0.64 | 0.88 |
| rs4430796 | A | G | 0.20 | 6/60/97 | 6/51/113 | 0.50 | 1.00 |
| rs4457053 | G | A | 0.26 | 11/64/87 | 14/61/94 | 1.00 | 0.43 |
| rs4607103 | T | C | 0.35 | 19/72/72 | 26/74/70 | 0.86 | 0.41 |
| rs4689388 | G | A | 0.30 | 11/76/76 | 9/81/79 | 0.20 | 0.06 |
| rs472265 | G | A | 0.39 | 21/85/56 | 20/91/59 | 0.25 | 0.11 |
| rs4812829 | A | G | 0.05 | 0/15/146 | 1/17/146 | 1.00 | 0.42 |
| rs564398 | C | T | 0.01 | 0/5/158 | 0/2/167 | 1.00 | 1.00 |
| rs5945326 | G | A | 0.12 | 2/27/99 | 1/20/65 | 1.00 | 1.00 |
| rs6467136 | A | G | 0.24 | 7/64/83 | 7/55/91 | 0.29 | 0.82 |
| rs6583826 | G | A | 0.48 | 28/86/42 | 38/89/37 | 0.20 | 0.35 |
| rs6815464 | G | C | 0.13 | 3/35/124 | 3/36/129 | 0.72 | 0.73 |
| rs7018475 | G | T | 0.12 | 1/37/123 | 2/33/130 | 0.47 | 1.00 |
| rs7041847 | G | A | 0.07 | 1/19/143 | 0/23/147 | 0.50 | 1.00 |
| rs7107217 | C | A | 0.47 | 39/73/46 | 33/90/45 | 0.34 | 0.36 |
| rs7172432 | A | G | 0.15 | 1/45/115 | 4/46/118 | 0.20 | 1.00 |
| rs7178572 | A | G | 0.40 | 27/82/54 | 22/86/62 | 0.75 | 0.42 |
| rs730570 | A | G | 0.20 | 5/58/98 | 6/53/107 | 0.48 | 1.00 |
| rs7542900 | T | C | 0.43 | 24/86/53 | 42/69/57 | 0.33 | 0.03 |
| rs7560163 | C | G | 0.20 | 11/41/109 | 7/52/109 | 0.02 | 0.81 |
| rs7578326 | G | A | 0.44 | 28/84/50 | 34/80/55 | 0.52 | 0.64 |
| rs7578597 | C | T | 0.24 | 8/67/88 | 7/60/102 | 0.41 | 0.82 |
| rs7630877 | A | G | 0.40 | 20/73/48 | 18/74/48 | 0.39 | 0.29 |
| rs7636 | A | G | 0.31 | 20/63/80 | 21/64/84 | 0.20 | 0.15 |
| rs7754840 | G | C | 0.43 | 19/87/28 | 12/88/40 | <0.001 | <0.001 |
| rs7903146 | T | C | 0.29 | 15/61/85 | 12/77/79 | 0.44 | 0.28 |
| rs7957197 | A | T | 0.28 | 14/67/82 | 10/73/87 | 1.00 | 0.34 |
| rs7961581 | C | T | 0.18 | 10/39/113 | 8/44/117 | 0.02 | 0.18 |
| rs8042680 | C | A | 0.01 | 0/5/158 | 0/3/167 | 1.00 | 1.00 |
| rs8050136 | A | C | 0.48 | 46/58/59 | 41/85/44 | <0.001 | 1.00 |
| rs831571 | T | C | 0.12 | 1/28/134 | 6/38/126 | 1.00 | 0.21 |
| rs864745 | C | T | 0.16 | 4/45/114 | 4/42/123 | 1.00 | 0.76 |
| rs896854 | C | T | 0.22 | 4/55/94 | 6/57/90 | 0.32 | 0.49 |
| rs9470794 | C | T | 0.26 | 13/58/92 | 13/65/92 | 0.41 | 0.70 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs9472138 | T | C | 0.31 | 11/68/84 | 17/82/69 | 0.70 | 0.39 |
| rs972283 | A | G | 0.11 | 2/37/124 | 2/27/141 | 1.00 | 0.63 |

MAF = Minor allele frequency; HWE = Hardy Weinberg Equilibrium; A = adenine; G = guanine; C = cytosine; T = thymine; X = minor allele; Y = major allele; Freq= frequency

## 4.5 GENETIC ASSOCIATION ANALYSIS

After the quality control procedures and the assessment of adherence to the assumptions of HWE as discussed above, 66 common variants were analysed via genetic association testing. The standard models of genetic association were used to evaluate the effect of the common variants on T2D risk, since the exact models of inheritance of all the SNPs were not reported. These genetic association models were the general genotypic, additive, dominant, recessive and CATT genetic models, which are described in detail in Section 3.9.3.

As elaborated on in the analysis strategy in Section 3.9.4, the differences in the genotypic distributions were first analysed using the univariate test (chi square test); the results are presented in Table 4.6. The multivariate test, that entailed the logistic regression test, was then conducted with age, sex and BMI as covariates. The results are presented in Table 4.7.

**Table 4.6 Genetic association results for the univariate tests of the selected SNPs**

| Gene | SNP | Minor Allele | Model | UNIVARIATE |
|---|---|---|---|---|
| | | | | EMP1 |
| **PSMD6** | rs831571 | T | ADD | 0.04 |
| | rs831571 | T | TREND | 0.03 |
| **FTO** | rs8050136 | A | DOM | 0.04 |
| | rs8050136 | A | GENO | 0.02 |
| **SLC44A3, F3** | rs7542900 | T | GENO | 0.03 |
| | rs7542900 | T | REC | 0.02 |
| **C2CD4B** | rs1436955 | C | TREND | 0.04 |

A1 = minor allele in the study sample. ADD = additive genetic model; DOM = dominant genetic model; GENO = general genetic model; TREND = Cochrane Armitage test of trend; REC = recessive genetic model; EMP1 = point-wise empirical value

Of the 66 common variants that were evaluated in the genetic association tests, only four SNPs, rs831571, rs8050136, rs7542900 and rs1436955, yielded significant results.

However, after correcting for multiple testing, none of the SNPs was noted to be significantly associated with IGT. Since the p-value is a test statistic for a single test, the multiple testing correction was done to control for the bias that can occur when multiple tests are conducted. However, it has also been reported that multiple testing can be too stringent, leading to loss of potential significant associations (Elsäßer *et al*., 2011). Therefore, the SNPs were regarded as having potential significant associations with T2D risk, which would, however, need to be validated through future well-powered genetic association studies.

The general genotypic and CATT models could not be evaluated by using logistic regression, as they have three genotype categories that are indicative of 2 degrees of freedom (df) variables. Logistic regression can only be used for categorical variables with two outcomes denoted as having 1 df, but cannot analyse categorical variables that have three outcomes depicted as 2 df (Purcell *et al.,* 2007). The different standard genetic models of disease penetrance have been reported to be related (Lewis & Knight, 2012). The additive models evaluate the distribution of alleles as described in Section 3.9.4, while the CATT model assesses the distributions of genotypes. However, although the CATT model evaluates genotypes and the additive genetic model assesses alleles, these models are similar, as they both evaluate how the increase in the order of the number of minor alleles is associated with disease risk (Lewis & Knight, 2012). Therefore a SNP can be statistically significant through the CATT and additive genotic model in the same population under study. The general genotypic genetic model does not involve any assumptions about the alleles and the genotypes of the SNPs. Therefore significant results may be obtained for this model, together with other models in the same population. This explains how a number of SNPs were noted to be significant in two different but related genetic association models in this study, as indicated in Table 4.6 and Table 4.7. The detailed results of the general genotypic, additive, dominant, recessive and the CATT genetic models are indicated in the appendix section.

**Table 4.7 Multivariate genetic association results for the selected SNPs**

| Gene | SNP | Minor Allele | Model | MULTIVARIATE | | |
|---|---|---|---|---|---|---|
| | | | | OR | EMP1 | EMP2 |
| **PSMD6** | rs831571 | T | ADD | 0.56 [0.34-0.92] | 0.03 | 0.76 |
| | rs831571 | T | DOM | 0.56 [0.34-0.92] | 0.04 | 0.94 |
| **FTO** | rs8050136 | A | DOM | 0.59 [0.36-0.97] | 0.04 | 0.90 |
| **SLC44A3, F3** | rs7542900 | T | REC | 0.51 [0.28-0.91] | 0.02 | 0.51 |
| **C2CD4B** | rs1436955 | C | ADD | 1.46 [1.04-2.04] | 0.03 | 0.84 |
| | rs1436955 | C | REC | 1.97 [1.06-3.65] | 0.03 | 0.65 |

A1 = minor allele in the study sample. ADD = additive genetic model; DOM = dominant genetic model; GENO = genotypic model; TREND = Cochrane Armitage test of trend; REC = recessive genetic mode (spelling mistake); OR = odds ratio; EMP1 = point-wise empirical value; EMP2 = family-wise correction for multiple comparisons.

The functional role of the SNPs that had significant associations was evaluated through the RegulomeDB database. The RegulomeDB database provides the functional predictions of the variants that are derived from GWASs (Boyle *et al*., 2012). All the SNPs that had significant associations were noted to have scores from 4-5 that were indicative of minimal binding evidence to transcription factors, except for rs831571, which had missing functional data. Most of the variants discovered through GWAS have been reported to be in non-coding regions and their role in disease aetiology is therefore postulated to be through the regulation of gene expression (Schaub *et al*., 2012). Therefore the binding to the transcription factors of these SNPs was evaluated as a regulatory proxy (Schaub *et al*., 2012). Thus it is probable that the variants that had significant associations in this study might have no known functional roles and might be in LD with a variant that has a functional role, since according to RegulomeDB they had minimal evidence of binding to the transcription factor. In subsequent sections the results of SNPs that had statistically significant results are discussed separately.

### 4.5.1 SNP rs831571

The *PSMD6* gene harbours the rs831571 SNP. The PSMD6 gene is involved in the degradation of ubiquitinated proteins (Lecker *et al*., 1999). This gene might be associated with T2D through the ubiquitin-proteasome system. The up-regulation of the ubiquitin-proteasome system in rat muscles was noted to be associated with T2D

(Lecker *et al*., 1999). However, limited evidence of the functional role of the SNP rs831571 was noted, as discussed in Section 4.5.1.

When the allele and genotype frequencies of the cases and controls were compared, significant differences were noted through the additive and CATT genetic models for this SNP, as indicated in Table 4.6. As discussed in Section 4.5, the CATT and additive genotypic models are related; therefore it is not unexpected that this SNP was noted to be significant through these two models. The significant association with this SNP implied that there were significant differences in the distribution of the alleles and genotypes, in the order of the increase in the number of the minor alleles carried by the cases and controls.

Multivariate analysis was then done to probe further into the extent to which the differences in the allelic and genotypic distributions of the SNP among the cases and controls were associated with IGT while adjusting for age, sex and BMI. The rs831571 SNP was discovered to have a protective association with IGT, as indicated by an OR of 0.56 through the additive ($p=0.03$) and the dominant genetic ($p=0.04$) models. However, after correcting for multiple testing, the association of rs831571 with IGT became statistically non-significant. In view of the implications of the multiple testing correction, this SNP was then regarded to have a potential protective association against T2D risk, as discussed in Section 4.5.

The CATT genetic model could not be evaluated using the logistic regression of having 2 df, as the logistic regression requires data to have only 1 df, as elaborated on in Section 4.5. Initially, through the univariate test, this SNP was noted to have significant results through the additive and CATT genetic models, but in the multivariate tests it was noted to have significant associations through the additive and the dominant genetic models. The discrepancy between the univariate and multivariate tests for this SNP might be because in the multivariate tests there was adjustment of other covariates, which might have caused this SNP to have significant associations through the dominant genetic model in which it was initially non-significant.

Through the additive model it could be deduced that the protective effect of rs831571 to IGT increased with the addition of each minor allele and through the dominant model it was noted that the protective effect of this SNP was due to the minor allele being dominant compared to the wild type allele. However, the protective effect of this SNP was contrary to what had been reported in earlier studies among the Asian population, where it was noted to increase T2D risk by an OR of 1.09 (Cho *et al.*, 2011). The discrepancy in the association patterns of the SNPs among black South Africans and the Asian population might be a result of the differences in the LD stretches among these populations (Teo *et al*., 2010). Since the SNPs that are derived from the association test are deemed to be in LD with the causal SNPs, the differences in LD may also imply that in the black South African population this SNP might be associated with a genetic variant that is protective against T2D and this might be contrary to the Asian population, where it is probably in LD with the variants that increase T2D risk (Teo *et al*., 2010). Therefore fine mapping strategies in this gene are required in the black South African population to identify the population-specific variants associated with increased risk of T2D.

### 4.5.2 SNP rs8050136

The *FTO* gene has been reported to be potentially involved in nucleic acid demethylation and the regulation of energy balance via the activity of the hypothalamus (Müller *et al*., 2013). Although genetic variation in this gene has been associated with T2D, there is still controversy about the role of the *FTO* gene in energy metabolism and how it affects obesity and T2D risk (Müller *et al*., 2013). Recent evidence has indicated that the obesity associated variants of the *FTO* gene to be linked with the expression of the homeobox gene *IRX3* and not the *FTO* gene in the human brain (Smemo *et al*., 2014). Thereby suggesting the *IRX3* to be associated with obesity through functional connections with the *FTO* gene variants. However, no evidence was noted for the functional role of the SNP rs8050136, as discussed in Section 4.5.1.

The rs8050136 SNP was noted to be significantly different among the cases and controls in the distribution of the genotypes through the general and dominant genetic models, as illustrated in Table 4.6. Since the general genetic model does not make any assumptions about the alleles in the SNP genotypes, these results implied that there

were significant differences in the genotypes of the cases and controls, without considering which of the alleles was dominant or recessive for this SNP, or whether there was a difference in the distribution of the genotypes as a result of the ordering of the genotypes according to the increase in the number of the minor alleles. The dominant genetic model compared the genotypes of the cases and controls of this SNP with the assumption that the minor allele was dominant. Therefore the significant results as determined when applying the dominant model for this SNP implied that there were significant differences in the distribution of the genotypes of the cases and controls among the carriers and non-carriers of the minor alleles of this SNP.

The associations of this SNP with IGT were further assessed through logistic regression while adjusting for age, sex and BMI. It was noted that this SNP was associated with a protective effect for T2D susceptibility, as indicated by an OR of 0.56 through the dominant genetic model ($p = 0.04$), as illustrated in Table 4.7. The significant results of the dominant genetic model revealed that the carriers of the minor allele of rs8050136 were protected against IGT compared to the non-carriers. Although the association of this SNP was noted as not being significantly associated after multiple testing correction, it was regarded to have a potential association with IGT, as discussed in Section 4.5.

A discrepancy was noted in the findings of this study compared to earlier studies with regard to the associations of rs8050136 SNP with increased T2D risk. Earlier studies reported that the rs8050136 SNP was associated with increased T2D risk among the European population with an OR of 1.23 (Zeggini *et al.,* 2007). The discrepancy of the findings of this study in respect of this SNP may be due to the differences in the LD patterns that have been noted to exist among the European and black South African populations (May *et al*., 2013; Teo *et al*., 2010). Therefore, as discussed in Section 4.5.1, the differences in the LD patterns may also be responsible for the discrepancy in the association patterns of this SNP among black South Africans and European populations.

### 4.5.3 SNP rs7542900

The rs7542900 SNP is located near the *SLC44A3* and *F3* genes. The SLC44A3 gene is involved in the transport of choline and the F3 gene, which is involved in coagulation (Griggs, 2007). Limited evidence exists of the role of these genes in T2D risk. However, limited evidence was discovered for the functional role of the SNP rs7542900, as discussed in Section 4.5.1.

The distribution of the genotypes of the cases and controls for the rs7542900 SNP was noted to be significantly different through the general and recessive genetic models, as illustrated in Table 4.6. The relationship between the recessive and general genotypic models is described in Section 4.5. However, these results implied that there were significant, general differences in the distribution of the genotypes among the cases and controls and that there were significant differences when the people who only carried the minor allele were compared with the carriers of the major allele among the cases and controls.

The differences in the genotypic distributions of this SNP among the cases and controls were further explored through logistic regression to evaluate to what extent they were associated with T2D risk while adjusting for age, sex and BMI. The rs7542900 SNP was noted to have a protective effect against T2D susceptibility as evidenced by an OR of 0.56 (p=0.02) through the recessive genetic model. This implied that in this study people who were homozygous for the minor allele were protected against T2D risk, as opposed to the carriers of the major allele. However, after correcting for multiple testing, the association of rs7542900 became statistically non-significant and it was then deemed to have a potential association with T2D risk, as discussed in Section in 4.5.

Contrary to this study, the rs7542900 SNP was reported to increase T2D risk, as indicated by an OR of 1.16 among African Americans (Palmer *et al*., 2012). The discrepancy in the association patterns of this study was also regarded as a result of the difference in LD patterns among African American populations and black South Africans, as discussed in Section 4.5.1 (May *et al.,* 2013).

### 4.5.4 SNP rs1436955

The rs1436955 SNP is located in the *C2CD4B* gene, which is involved in the regulation of cell architecture and adhesion (Warton *et al*., 2004). This gene has also been noted to be potentially involved in inflammation (Warton *et al*., 2004). Although variants in this gene have been associated with T2D, little information exists about its role in T2D. However, limited evidence of the functional role of the SNP rs1387153 was discovered, as discussed in Section 4.5.1.

Initially, the genotype distributions of the cases and controls were compared for this SNP via the univariate tests and significant differences were noted through the CATT genetic model. These results implied that there were significant differences in the order of the increase in the number of the minor alleles carried by the cases and controls, as illustrated in Table 4.6. Through logistic regression the association of this SNP with IGT was further explored while adjusting for age, sex and BMI. The SNP rs1436955 was the only variant noted to increase the risk of IGT with an odds ratio of 1.46 and 1.96, using the additive and recessive genetic models of disease penetrance in this study. Although the association of this SNP was non-significant after multiple testing correction, it was regarded that this SNP might have potential association with increased T2D risk, as elaborated on in Section 4.5.

The significant results through the additive genetic model implied that the SNP increased T2D risk in relation to the number of minor alleles a person had. The significant results as determined via the recessive genetic model also implied that this SNP had a significant effect among people who were homozygous for it, compared to those who were carriers of the major allele. The additive genetic model of genetic penetrance is evaluated mostly in GWAS studies, since it also accounts for the dominant model effect (Bush *et al*., 2012). However, the additive genetic model has been noted to be under-powered when used to evaluate the recessive model of disease penetrance effects (Lettre *et al*., 2007). This might therefore be the reason for the differences in the OR for the recessive and additive model of disease penetrance noted for this SNP. However, in order to compare the results of this study with other studies, the results of the additive model, which is commonly used in GWAS studies, were used.

Earlier studies reported that the SNP rs1436955 was associated with an increase in the risk of T2D among Asian individuals with an odds ratio of 1.13 (Shu *et al*., 2010). Although the odds ratio reported for this SNP among Asians was different from that in this study, both studies had a similar pattern of association, which indicated that rs1436955 increased the risk of T2D.

### 4.5.5 Possible reasons for discrepancy of association patterns of selected common variants

Although evidence exists for the sharing of common variants that increase T2D risk across different ethnicities, only one variant, rs1436955, was noted to increase the risk of T2D in this study from among the 66 common variants reported to increase T2D risk in other ethnicities (Sanghera & Blackett, 2012). Potential reasons for the lack and discrepancy of significant associations of the other SNPs that were noted to be non-significant in this study are discussed below.

#### 4.5.5.1 Low power of study

The sample size of this study had low power to detect SNPs with modest associations with T2D risk. According to the sample size calculations elaborated on in Chapter 3, the sample size was able to detect variants with an odds ratio greater than 1.65 and a frequency greater than 20%. However, most of the SNPs associated with T2D were noted to have odds ratios between 1.2 and 1.3 and a sample size of 1 146 cases and controls is required to have 80% power to detect such SNPs at a significance level of 0.05 (Purcell *et al*., 2003). Other studies in Africa that have been able to replicate T2D risk-associated variants have been reported to be well-powered. For instance, the study by Cauchi *et al*. (2012) was well-powered with a sample size of 2000 cases and controls and it discovered that 13 common variants associated with T2D were shared among Europeans and North African Arabs from Tunisia and Morocco. However, in view of the challenges of conducting the OGTT, the sample size of 360 was justifiable and considered for use in this study. In addition, a similar sample size had been used to evaluate T2D common variants in a black South African population of Zulu descent and that study had reported significant associations, which then became a motivation to use a similar sample size. The data from this study can be included in future T2D meta-

analysis genetic association studies to improve the power of detecting genetic variants associated with T2D risk.

#### 4.5.5.2 Genetic diversity of black South Africans compared to other ethnicities

GWASs rely on assessing common genetic variation and the manner in which it relates to disease susceptibility by using tagging SNPs, which are chosen based on the data available from the International HapMap Project (Bush & Moore, 2012). These tagging SNPs are not actually the causal variants but are in LD with the causal variants (Bush & Moore, 2012). Although African populations were incorporated in the HapMap project, evidence now exists that black South Africans have distinct genetic variation patterns from other ethnicities who were part of the HapMap project (May *et al*., 2013). It might thus be plausible that the selected tagging SNPs associated with T2D in the GWAS might not have been representative of the genetic variation in black South Africans and hence few of these variants were replicated in this study. In addition to this, since the tagging SNPs are in LD with causal variants, the differences in LD across populations might have had an effect on the replication of the genetic associations. The LD stretches in African populations have been reported to be shorter compared to Caucasian populations, such that replication of genetic associations is less probable, unless the causal variant is actually included in the genetic association (Teo *et al*., 2010).

However, some studies done in Africa have been able to replicate common variants associated with T2D that were reported in European populations. Thirteen variants associated with T2D risk were reported to be shared among North African Arabs from Tunisia and Morocco and European populations. Such evidence would suggest that black South Africans might also share significant association with other ethnicities. However, evidence from the analysis of the mitochondrial DNA of the North African and European populations indicated that these two populations shared common haplogroups, which might have led to them also having shared common variants associated with T2D risk (Ennaffa *et al*., 2009). However, it has been reported that black South Africans have genetic variations that are different from those of North African Arabs and European populations (May *et al*., 2013; Teo *et al*., 2010). Therefore

it can be expected that black South Africans might not share common variants associated with T2D with other ethnicities, as documented in this study.

#### 4.5.5.3 Intra-genetic diversity

The SNP rs7903146 of the *TCF7L2* gene was reported to be associated with T2D risk in black South Africans of Zulu descent (Pirie *et al*., 2010). However, this association was not replicated in this study, regardless of the fact that the study by Pirie et al (2010) had a sample size of 178 cases and 200 controls that was similar to that of this study. Evidence exists for a high level of genetic diversity that may exist among different population groups in Africa and this might have led to the discrepancy in the association pattern of rs7903145 among the Zulu and Tswana populations (Ramsay, 2012). However, more studies are required to validate whether the Zulu and Tswana are genetically diverse, as they have been noted to have originated from the Bantu-speaking population group (May *et al*., 2013).

## 4.6 SUMMARY OF THE RESULTS

The results of this study will be summarised by discussing how the objectives indicated in Chapter 1 were met. The first objective was to determine the risk factors for T2D (BMI, dietary fat intake and dietary carbohydrate intake) profile of the cases and controls to adjust for confounding variables. BMI was noted to be statistically different among the cases and controls and therefore it was controlled for in the genetic association tests.

The second objective was to determine the allelic and genotypic distributions of the selected SNPs among individuals with IGT and control individuals in a black South African population. The genotypic distributions were assessed for their adherence to the assumptions of HWE. Three SNPs, which included rs7754840 ($p < 0.001$), rs13266634 ($p < 0.001$) and rs7542900 ($p = 0.03$), did not adhere to the assumptions of HWE among the controls, whereas among the cases, the genotypic distributions of the following SNPs were noted to deviate from the HWE assumptions: rs8050136 ($p < 0.001$), rs7754840 ($p < 0.001$), rs7961581 ($p = 0.02$) and rs7560163 ($p = 0.02$). The SNPs rs7542900, rs7560163 and rs7961581 were retained for the genetic association

analysis on the basis of having acceptable SNP genotype clusters, which ruled out the possibility of poor genotyping. The departures from HWE were noted to be a probable result of poor genotyping of the homologous genome region for rs7754840.

The third objective was to compare the association patterns of the statistically significant SNPs in the study sample to those reported from other ethnicities. After correcting for multiple testing, none of the SNPs indicated a significant association with IGT. However, five SNPs that will need to be evaluated in a larger sample were noted to have significant associations with IGT. Of the five SNPs only one, rs1436955, had a similar association pattern to what had been reported in other ethnicities. The other SNPs, which consisted of rs8050136, rs7542900, rs1387153 and rs831571, were noted to be protective rather than to increase T2D risk. The differences in the LD patterns of black South Africans and Europeans and other ethnicities were deemed to have led to the discrepancy in association patterns. The SNPs that were evaluated are presumed not to be the causal SNPs but to be in LD with them, therefore the SNPs that had protective associations might have been in LD with SNPs that were protective against T2D risk, unlike in other ethnicities where they were initially reported to be associated with increased T2D risk.

# CHAPTER 5

# CONCLUSION

The continual increase in the prevalence of T2D globally is now a cause of concern. Limited information about the aetiology of T2D exists, particularly the precise genetic factors that increase T2D risk. Little has been reported about the common variants associated with T2D risk in the black South African population. To date, 79 common variants have been associated with T2D in other ethnicities, but the majority of these SNPs have not been evaluated in the black South African population. Evidence exists for the sharing of common variants that are associated with T2D among different ethnicities and this was the motivation used to evaluate these common variants in this study in the black South African population. This chapter elaborates on the summary of the main findings, the implications of the results, the limitations of the methodology that was used and recommendations for future studies.

## 5.1 SUMMARY OF THE MAIN FINDINGS

Of the 79 common variants currently associated with T2D, only 66 could be evaluated through the genetic association testing process implemented in this study after the quality control procedures had been undertaken, as indicated in Section 4.5. Four SNPs among the 66 SNPs, which were genotyped successfully in this study, were significantly associated with T2D risk, as indicated by IGT. The four common variants included rs831571, rs8050136, rs7542900 and rs1436955, which are located in the *PSMD6, FTO, SLC44A3/F3* and *C2CD4B* genes respectively. Of these four variants, only rs1436955 was associated with increased risk of T2D, while the other variants, which included rs831571, rs8050136 and rs7542900, were associated with reduced risk of T2D. The protective nature of the variants was hypothesised to be a result of the differences in the LD patterns among black South Africans and other ethnicities. Since the associated SNPs are likely not to be the causal variants, the differences in LD in the African population may have led to the investigated SNPs being in LD with other variants, which are protective against T2D risk, as opposed to those that increase T2D risk. This is also a possible reason why most of the SNPs presented with no association

to T2D disease risk, although the limitations, as discussed in Section 5.3, may also have contributed to this.

## 5.2 IMPLICATIONS OF THIS STUDY

Although evidence for shared genetic variants associated with T2D risk exists among North African Arabs, Europeans, Mexicans and Asians, this may not be the case for the black South African population (Waters *et al*., 2010; Cauchi *et al*., 2012). It is probable that the common variants associated with T2D risk in black South Africans are different from those that have been reported in other ethnicities. Common variant association testing is strongly dependent upon LD and earlier studies have reported that the LD patterns of people in Africa are more complex and thus different from other ethnicities (Teo *et al*., 2010). Therefore the results of this study add to the body of knowledge, which indicates that common variants discovered in non-African countries cannot be used to infer the genetic variants that cause disease susceptibility in African populations, especially in Southern Africa (May *et al*., 2013). However, there is also evidence of high levels of genetic diversity among people in Southern Africa, which might have led to the failure to replicate the association of rs7903146 among people of Tswana origin that had previously been reported among people of Zulu descent (Teo *et al*., 2010; Ramsay, 2011).

The clinical utility of using the common variants for direct-to-consumer genetic testing has recently been questioned by the Southern African Society of Human Genetics (SASHG) for complex diseases such as T2D in view of the limited informativeness of these variants in non-African countries with regard to T2D heritability (Dandara *et al*., 2013). Since the majority of the SNPs that have been associated with T2D in other ethnicities were noted not to be associated with T2D risk in this black South African population of Tswana descent, this supports the stance of the SASHG against direct-to-consumer genetic testing for South Africans in general. Although more studies might be required to validate the findings from this research, it is pivotal for local health authorities to protect the general public through the provision of evidence-based policies regarding direct-to-consumer genetics tests.

## 5.3 LIMITATIONS

A few limitations were noted in this study. Although the sample size was justified by being comparable with other local studies that had replicated GWAS common variant associations with T2D among black South Africans, there is evidence that sample sizes of about 1 000 are better able to replicate GWAS common variants (Purcell *et al.,* 2003). This might have contributed to the replication of the variants associated with T2D in other ethnicities in this study. During the pre-assessment analysis of the genotyping of the SNPs, two SNPs had poor metrics for successful genotyping and they were dropped from the list. It is possible that these two SNPs might have yielded significant associations with IGT if they had been analysed. Furthermore, the 11 SNPs that were not included in the genetic analysis because their genotyping metrics were indicative of poor genotyping with the BeadXpress® reader might also have indicated significant associations. However, since most of the variants in this study did not present with any significant associations, it is unlikely that this would have been the case.

## 5.4 RECOMMENDATIONS

Well-powered genetic association studies are required to evaluate the association of genetic variation with complex disease phenotypes among black South Africans. The Southern African Human Genome Programme (SAHGP) is anticipated to unravel the genetic variation in Southern Africa and thus enable the discovery of common genetic variants that are better suited for genetic association testing for T2D risk in black South Africans (May *et al*., 2013). Genetic susceptibility to T2D may also be due to rare variants and/or CNVs that were not analysed in this study (Craddock *et al*., 2010). Therefore there is a great need to sequence the genes that harbour these common variants, as these are hypothesised to have rare variants as well, which might have greater effects on T2D risk (Zeggini *et al*., 2008). Next generation sequencing might also help with the identification of population-specific variants associated with T2D risk (May *et al*., 2013). Evaluation of the role of CNVs in T2D genetic susceptibility in black South Africans is also required (Groop & Pociot, 2013). There is a need to explore the functional role of the variants associated with T2D as well to exploit the clinical utility of

the genetic variants in treatment (pharmacogenomics) and preventative interventions to curb the increase in T2D (Groop & Pociot, 2013). T2D is hypothesised to develop in response to the interaction of both genetic and environmental factors (Goh, 2012). However, information is still limited regarding the extent of this interaction in the black South African population. Overall, a systems biological approach will be required to fully understand the manner in which the variants lead to the development of T2D and this will open avenues for many ways of curtailing the global increase in T2D through relevant preventative and curative interventions (Franks *et al*., 2013).

## 5.5 CONCLUSION

This study was a unique investigation that set out to evaluate the majority of the previously reported common variants associated with T2D from other ethnicities in a black South African population. However, according to the findings of this study it is highly probable that the black South African population share very few common variants associated with T2D with other ethnicities. This study supports what has been reported in a number of local genetics studies in South Africa, namely that the genetic background of black South Africans is diverse and cannot be evaluated using variants from other ethnicities. The results of this study also support the stance taken by local genetic societies to discourage direct-to-consumer genetic tests for complex diseases such as T2D, as these tests rely on common variants reported in other ethnicities, which were noted not be associated with T2D risk among black South Africans of Tswana descent. The common variants that have been reported to be associated with T2D in other ethnicities might currently also not be of significant clinical utility even to be used in pharmacogenomics analyses to personalise treatment of T2D among black South Africans. More genetic studies are thus required to elucidate the genetic susceptibility to T2D risk in the black South African population in order to understand T2D aetiology better and develop relevant preventative and curative interventions to curb the T2D epidemic. In view of the results of this study, the current initiatives that have been set up to evaluate genetic susceptibility to T2D in Africa, such as the H3 Africa programme, should focus on identifying population-specific genetic variation using next-generation sequencing. Projects such as the SAHGP, which are working on documenting the genetic variation in Southern Africa, should be supported to enable the use of more

representative tagging SNPs, which could be used in future GWAS studies to identify common variants associated with T2D risk. Therefore the results of this study highlight the need for population specific-variants to evaluate T2D genetic susceptibility in the black South African population.

# REFERENCES

Anon. Preventing chronic diseases. A vital investment: WHO global report. Geneva: World Health Organization, 2005. pp. 200. CHF 30.00.http://ije.oxfordjournals.org.nwulib.nwu.ac.za/content/35/4/1107.full Date of access: 8/12/2013.

1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. & McVean, G.A. 2010. A map of human genome variation from population-scale sequencing. *Nature,* 467(7319):1061-1073.

Ahlqvist, E., Ahluwalia, T.S. & Groop, L. 2011. Genetics of type 2 diabetes. *Clinical chemistry,* 57(2):241-254.

Alberti, K.G. & Zimmet, P.Z. 1998. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabetic medicine: A journal of the British diabetic association,* 15(7):539-553.

Alberti, K.G.M.M. 1996. The clinical implications of impaired glucose tolerance. *Diabetic medicine,* 13(11):927-937.

Alonso-Magdalena, P., Morimoto, S., Ripoll, C., Fuentes, E. & Nadal, A. 2006. The estrogenic effect of bisphenol A disrupts pancreatic ß-cell function in vivo and induces insulin resistance. *Environmental health perspectives,* 114(1):106-112.

Amod A, Ascott-Evans BH, Berg GI, Blom DJ, Brown SL, Carrihill MM, Dave JA, Distiller LA, Ganie YN, Grobler N, Heilbrunn AG, Huddle KRL, Janse van Rensburg G, Jivan D, Joshi P, Khutsoane DT, Levitt NS, May WM, Mollentze WF, Motala AA, Paruk IM, Pirie FJ, Raal FJ, Rauff S, Raubenheimer PJ, Randeree HAR, Rheeder P, Tudhope L, Van Zyl DJ, Young M & Guideline Committee. 2012. The 2012 SEMDSA guideline for the management of type 2 diabetes (revised). *Journal of endocrinology, metabolism and diabetes in South Africa,* 17(1):S1-S94.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T. 2010. Data quality control in genetic case-control association studies. *Nature protocols,* 5(9):1564-1573.

Barker, D.J. 2007. The origins of the developmental origins theory. *Journal of internal medicine,* 261(5):412-417.

Barr, E.L., Zimmet, P.Z., Welborn, T.A., Jolley, D., Magliano, D.J., Dunstan, D.W., Cameron, A.J., Dwyer, T., Taylor, H.R., Tonkin, A.M., Wong, T.Y., McNeil, J. & Shaw, J.E. 2007. Risk of cardiovascular and all-cause mortality in individuals with diabetes mellitus, impaired fasting glucose, and impaired glucose tolerance: The Australian diabetes, obesity, and lifestyle study (AusDiab) *Circulation,* 116(2):151-157.

Barr, R.G., Nathan, D.M., Meigs, J.B. & Singer, D.E. 2002. Tests of glycemia for the diagnosis of type 2 diabetes mellitus. *Annals of internal medicine,* 137(4):263-272.

Barroso, I., Luan, J., Middelberg, R.P., Harding, A.H., Franks, P.W., Jakes, R.W., Clayton, D., Schafer, A.J., O'Rahilly, S. & Wareham, N.J. 2003. Candidate gene association study in type 2 diabetes indicates a role for genes involved in beta-cell function as well as insulin action. *PLoS biology,* 1(1):E20.

Bergholdt, R., Brorsson, C., Palleja, A., Berchtold, L.A., Floyel, T., Bang-Berthelsen, C.H., Frederiksen, K.S., Jensen, L.J., Storling, J. & Pociot, F. 2012. Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression. *Diabetes,* 61(4):954-962.

Betterle, C., Zanette, F., Pedini, B., Presotto, F., Rapp, L.B., Monciotti, C.M. & Rigon, F. 1984. Clinical and subclinical organ-specific autoimmune manifestations in type 1 (insulin-dependent) diabetic patients and their first-degree relatives. *Diabetologia,* 26(6):431-436.

Bodmer, W. 1999. Familial adenomatous polyposis (FAP) and its gene, APC. *Cytogenetics and cell genetics,* 86(2):99-104.

Bollati, V. & Baccarelli, A. 2010. Environmental epigenetics. *Heredity,* 105(1):105-112.

Bonnefond, A., Clément, N., Fawcett, K., Yengo, L., Vaillant, E., Guillaume, J., Dechaume, A., Payne, F., Roussel, R. & Czernichow, S. 2012. Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nature genetics,* 44(3):297-301.

Bouatia-Naji, N., Bonnefond, A., Cavalcanti-Proença, C., Sparso, T., Holmkvist, J., Marchand, M., Delplanque, J., Lobbens, S., Rocheleau, G., Durand, E., De Graeve, F., Chèvre, J.-., Borch-Johnsen, K., Hartikainen, A.., Ruokonen, A., Tichet, J., Marre, M., Weill, J., Heude, B., Tauber, M., Lemaire, K., Schuit, F., Elliott, P., Jørgensen, T., Charpentier, G., Hadjadj, S., Cauchi, S., Vaxillaire, M., Sladek, R., Visvikis-Siest, S., Balkau, B., Lévy-Marchal, C., Pattou, F., Meyre, D., Blakemore, A.I.F., Jarvelin, M.-., Walley, A.J., Hansen, T., Dina, C., Pedersen, O. & Froguel, P. 2009. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nature genetics,* 41(1):89-94.

Bradshaw, D., Norman, R., Pieterse, D., Levitt, N.S. & South African Comparative Risk Assessment Collaborating Group. 2007. Estimating the burden of disease attributable to diabetes in South Africa in 2000 *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde,* 97(8 Pt 2):700-706.

Brasacchio, D., Okabe, J., Tikellis, C., Balcerczyk, A., George, P., Baker, E.K., Calkin, A.C., Brownlee, M., Cooper, M.E. & El-Osta, A. 2009. Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase enzymes associated with gene-activating epigenetic marks that coexist on the lysine tail. *Diabetes,* 58(5):1229-1236.

Bush, W.S. & Moore, J.H. 2012. Chapter 11: Genome-wide association studies. *PLoS computational biology,* 8(12):e1002822.

Campbell, P.J. & Carlson, M.G. 1993. Impact of obesity on insulin action in NIDDM. *Diabetes,* 42(3):405-410.

Cauchi, S., Ezzidi, I., El Achhab, Y., Mtiraoui, N., Chaieb, L., Salah, D., Nejjari, C., Labrune, Y., Yengo, L. & Beury, D. 2012. European genetic variants associated with type 2 diabetes in North African Arabs. *Diabetes & metabolism,* 38(4):316-323.

Chen, G., Adeyemo, A., Zhou, J., Chen, Y., Huang, H., Doumatey, A., Lashley, K., Agyenim-Boateng, K., Eghan, B.A.,Jr, Acheampong, J., Fasanmade, O., Johnson, T., Okafor, G., Oli, J., Amoah, A. & Rotimi, C. 2007. Genome-wide search for susceptibility genes to type 2 diabetes in West Africans: Potential role of C-peptide. *Diabetes research and clinical practice,* 78(3):e1-6.

Cheng, C., Kushner, H. & Falkner, B.E. 2006. The utility of fasting glucose for detection of prediabetes. *Metabolism: Clinical and experimental,* 55(4):434-438.

Cho, Y.S., Chen, C.H., Hu, C., Long, J., Ong, R.T., Sim, X., Takeuchi, F., Wu, Y., Go, M.J., Yamauchi, T., Chang, Y.C., Kwak, S.H., Ma, R.C., Yamamoto, K., Adair, L.S., Aung, T., Cai, Q., Chang, L.C., Chen, Y.T., Gao, Y., Hu, F.B., Kim, H.L., Kim, S., Kim, Y.J., Lee, J.J., Lee, N.R., Li, Y., Liu, J.J., Lu, W., Nakamura, J., Nakashima, E., Ng, D.P., Tay, W.T., Tsai, F.J., Wong, T.Y., Yokota, M., Zheng, W., Zhang, R., Wang, C., So, W.Y., Ohnaka, K., Ikegami, H., Hara, K., Cho, Y.M., Cho, N.H., Chang, T.J., Bao, Y., Hedman, A.K., Morris, A.P., McCarthy, M.I., DIAGRAM Consortium, MuTHER Consortium, Takayanagi, R., Park, K.S., Jia, W., Chuang, L.M., Chan, J.C., Maeda, S., Kadowaki, T., Lee, J.Y., Wu, J.Y., Teo, Y.Y., Tai, E.S., Shu, X.O., Mohlke, K.L., Kato, N., Han, B.G. & Seielstad, M. 2011. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in East Asians. *Nature genetics,* 44(1):67-72.

Chong, S. & Whitelaw, E. 2004. Epigenetic germline inheritance. *Current opinion in genetics and development,* 14(6):692-696.

Churchill, G.A. 2013. Epistasis. (In Maloy, S. & Hughes, K., eds. Brenner's encyclopedia of genetics (second edition). Academic Press. p. 505-507).

Clarke, G.M., Anderson, C.A., Pettersson, F.H., Cardon, L.R., Morris, A.P. & Zondervan, K.T. 2011. Basic statistical analysis in genetic case-control studies. *Nature protocols,* 6(2):121-133.

Conway, A., Rolley, J.X., Fulbrook, P., Page, K. & Thompson, D.R. 2013. Improving statistical analysis of matched case-control studies. *Research in nursing & health,* 36(3):320-324.

Cox, M.E. & Edelman, D. 2009. Tests for screening and diagnosis of type 2 diabetes. *Clinical diabetes,* 27(4):132-138.

Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F. & Giannoulatou, E. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature,* 464(7289):713-720.

Dabelea, D. & Pettitt, D.J. 2001. Intrauterine diabetic environment confers risks for type 2 diabetes mellitus and obesity in the offspring, in addition to genetic susceptibility. *Journal of pediatric endocrinology and metabolism,* 14(8):1085-1091

Dandara, C., Greenberg, J., Lambie, L., Lombard, Z., Naicker, T., Ramesar, R., Ramsay, M., Roberts, L., Theron, M. & Venter, P. 2013. Direct-to-consumer genetic testing: To test or not to test, that is the question. *SAMJ: South African medical journal,* 103(8):510-512.

Davidson, M.B., Schriger, D.L., Peters, A.L. & Lorber, B. 1999. Relationship between fasting plasma glucose and glycosylated hemoglobin: Potential for false-positive diagnoses of type 2 diabetes using new diagnostic criteria. *Journal of the American medical association,* 281(13):1203-1210.

Davies, M.J., Rayman, G., Gray, I.P., Day, J.L. & Hales, C.N. 1993. Insulin deficiency and increased plasma concentration of intact and 32/33 split proinsulin in subjects with impaired glucose tolerance. *Diabetic medicine,* 10(4):313-320.

Diamanti-Kandarakis, E., Bourguignon, J., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T. & Gore, A.C. 2009. Endocrine-disrupting chemicals: An endocrine society scientific statement. *Endocrine reviews,* 30(4):293-342.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., Lindgren, C.M., Mägi, R., Morris, A.P., Randall, J., Johnson, T., Elliott, P., Rybin, D., Thorleifsson, G., Steinthorsdottir, V., Henneman, P., Grallert, H., Dehghan, A., JanHottenga, J., Franklin, C.S., Navarro, P., Song, K., Goel, A., Perry, J.R.B., Egan, J.M., Lajunen, T., Grarup, N., Sparsø, T., Doney, A., Voight, B.F., Stringham, H.M., Li, M., Kanoni, S., Shrader, P., Cavalcanti-Proença, C., Kumari, M., Qi, L., Timpson, N.J., Gieger, C., Zabena, C., Rocheleau, G., Ingelsson, E., An, P., O'Connell, J., Luan, J., Elliott, A., McCarroll, S.A., Payne, F., Roccasecca, R.M., Pattou, F., Sethupathy, P., Ardlie, K., Ariyurek, Y., Balkau, B., Barter, P., Beilby, J.P., Ben-Shlomo, Y., Benediktsson, R., Bennett, A.J., Bergmann, S., Bochud, M., Boerwinkle, E., Bonnefond, A., Bonnycastle, L.L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Bumpstead, S.J., Charpentier, G., Der IdaChen, Y., Chines, P., Clarke, R., McOin, L.J., Cooper, M.N., Cornelis, M., Crawford, G., Crisponi, L., Day, I.N.M., De Geus, E.J.C., Delplanque, J., Dina, C., Erdos, M.R., Fedson, A.C., Fischer-Rosinsky, A., Forouhi, N.G., Fox, C.S., Frants, R., GraziaFranzosi, M., Galan, P., Goodarzi, M.O., Graessler, J., Groves, C.J., Grundy, S., Gwilliam, R., Gyllensten, U., Hadjadj, S., Hallmans, G., Hammond, N., Han, X., -LiisaHartikainen, A., Hassanali, N., Hayward, C., Heath, S.C., Hercberg, S., Herder, C., Hicks, A.A., Hillman, D.R., Hingorani, A.D., Hofman, A., Hui, J., Hung, J., Isomaa, B., Johnson, P.R.V., Jørgensen, T., Jula, A., Kaakinen, M., Kaprio, J., AnteroKesaniemi, Y., Kivimaki, M., Knight, B., Koskinen, S., Kovacs, P., Kyvik, K.O., Lathrop, G.M., Lawlor, D.A., Bacquer, O.L., Lecoeur, C., Li, Y., Lyssenko, V., Mahley, R., Mangino, M., Manning, A.K., TeresaMartínez-Larrad, M., McAteer, J.B., McCulloch, L.J., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B.D., Morken, M.A., Mukherjee, S., Naitza, S., Narisu, N., Neville, M.J., Oostra, B.A., Orrù, M., Pakyz, R., APalmer, C.N.,

Paolisso, G., Pattaro, C., Pearson, D., Peden, J.F., Pedersen, N.L., Perola, M., Pfeiffer, A.F.H., Pichler, I., Polasek, O., Posthuma, D., Potter, S.C., Pouta, A., Province, M.A., Psaty, B.M., Rathmann, W., Rayner, N.W., Rice, K., Ripatti, S., Rivadeneira, F., Roden, M., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A.A., Scheet, P., Scott, L.J., Seedorf, U., Sharp, S.J., Shields, B., Sigursson, G., Sijbrands, E.J.G., Silveira, A., Simpson, L., Singleton, A., Smith, N.L., Sovio, U., Swift, A., Syddall, H., -ChristineSyvänen, A., Tanaka, T., Thorand, B., Tichet, J., Tönjes, A., Tuomi, T., Uitterlinden, A.G., Van Dijk, K.W., Hoek, M.V., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P.J., Walley, A., BragiWalters, G., Ward, K.L., Watkins, H., Weedon, M.N., Wild, S.H., Willemsen, G.M., Witteman, J.C., GYarnell, J.W., Zeggini, E., Zelenika, D., Zethelius, B., Zhai, G., Zhao, J.H., Zillikens, M.C., Borecki, I.B., Loos, R.J.F., Meneton, P., Magnusson, P.K.E., Nathan, D.M., Williams, G.H., Hattersley, A.T., Silander, K., Salomaa, V., Smith, G.D., Bornstein, S.R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A.R., Cooper, C., Dedoussis, G.V., Serrano-Ríos, M., Morris, A.D., Lind, L., Palmer, L.J., Hu, F.B., Franks, P.W., Ebrahim, S., Marmot, M., Linda Kao, W.H., Pankow, J.S., Sampson, M.J., Kuusisto, J., Laakso, M., Hansen, T., Pedersen, O., Pramstaller, P.P., Wichmann, H.E., Illig, T., Rudan, I., Wright, A.F., Stumvoll, M., Campbell, H., Wilson, J.F., Bergman, R.N., Buchanan, T.A., Collins, F.S., Mohlke, K.L., Tuomilehto, J., Valle, T.T., Altshuler, D., Rotter, J.I., Siscovick, D.S., Penninx, B.W.J.H., Boomsma, D.I., Deloukas, P., Spector, T.D., Frayling, T.M., Ferrucci, L., Kong, A., Thorsteinsdottir, U., Stefansson, K., Van Duijn, C.M., Aulchenko, Y.S., Cao, A., Scuteri, A., Schlessinger, D., Uda, M., Ruokonen, A., -RiittaJarvelin, M., Waterworth, D.M., Vollenweider, P., Peltonen, L., Mooser, V., Abecasis, G.R., Wareham, N.J., Sladek, R., Froguel, P., Watanabe, R.M., Meigs, J.B., Groop, L., Boehnke, M., McCarthy, M.I., Florez, J.C. & Barroso, I. 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature genetics,* 42(2):105-116.

Elbein, S.C. 1997. The genetics of human noninsulin-dependent (type 2) diabetes mellitus. *The journal of nutrition,* 127(9):1891S-1896S.

Elsäßer, A., Victor, A. & Hommel, G. 2011. Multiple testing in candidate gene situations: A comparison of classical, discrete, and resampling-based procedures. *Statistical applications in genetics and molecular biology,* 10e51.

Ennafaa, H., Cabrera, V.M., Abu-Amero, K.K., Gonzalez, A.M., Amor, M.B., Bouhaha, R., Dzimiri, N., Elgaaied, A.B. & Larruga, J.M. 2009. Mitochondrial DNA haplogroup H structure in North Africa. *BMC genetics,* 108-2156-10-8.

Evans, D.M., Marchini, J., Morris, A.P. & Cardon, L.R. 2006. Two-stage two-locus models in genome-wide association. *PLoS genetics,* 2(9), e157.

Ferrell, R.E. & Lyengar, S. 1993. Molecular studies of the genetics of non-insulin-dependent diabetes mellitus *American journal of human biology,* 5(4):415-424.

Florez, J.C. 2013. Chapter 82 - diabetes. (In Ginsburg, G.S. & Willard, H.F., eds. Genomic and personalized medicine (second edition). Academic Press. p. 990-1005).

Franks, P.W., Looker, H.C., Kobes, S., Touger, L., Tataranni, P.A., Hanson, R.L. & Knowler, W.C. 2006. Gestational glucose tolerance and risk of type 2 diabetes in young pima indian offspring. *Diabetes,* 55(2):460-465.

Franks, P.W. 2011. Gene x environment interactions in type 2 diabetes. *Current diabetes reports,* 11(6):552-561.

Franks, P.W., Pearson, E. & Florez, J.C. 2013. Gene-environment and gene-treatment interactions in type 2 diabetes: Progress, pitfalls, and prospects. *Diabetes care,* 36(5):1413-1421.

George, K. & Alberti, M.M. 1998. Impaired glucose tolerance: What are the clinical implications? *Diabetes research and clinical practice,* 40:S3 - S8.

Gerken, T., Girard, C.A., Tung, Y.C., Webby, C.J., Saudek, V., Hewitson, K.S., Yeo, G.S., McDonough, M.A., Cunliffe, S., McNeill, L.A., Galvanovskis, J., Rorsman, P.,

Robins, P., Prieur, X., Coll, A.P., Ma, M., Jovanovic, Z., Farooqi, I.S., Sedgwick, B., Barroso, I., Lindahl, T., Ponting, C.P., Ashcroft, F.M., O'Rahilly, S. & Schofield, C.J. 2007. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science (New York, N.Y.),* 318(5855):1469-1472.

Goh, K. 2012. Chapter 17 - Gene–environment interaction in the pathogenesis of type 2 diabetes. (In Bagchi, D. & Sreejayan, N., eds. Nutritional and therapeutic interventions for diabetes and metabolic syndrome. Academic Press. p. 211-224).

Greiner, T. & Bäckhed, F. 2011. Effects of the gut microbiota on obesity and glucose homeostasis. *Trends in endocrinology and metabolism,* 22(4):117-123.

Griggs, B.L. 2007. Functional characterization of ectodysplasin-1 and identification of novel candidate genes for ectodermal dysplasias and associated clinical features. Clemson University. (Thesis-PhD).

Groop, L. & Pociot, F. 2014. Genetics of diabetes – Are we missing the genes or the disease? *Molecular and cellular endocrinology,* 382(1):726-739.

Hardison, R.C. 2012. Genome-wide epigenetic data facilitate understanding of disease susceptibility association studies. *The journal of biological chemistry,* 287(37):30932-30940.

Harris, M.I. 1993. Undiagnosed NIDDM: Clinical and public health issues *Diabetes care,* 16(4):642-652.

Hirschhorn, J.N. & Daly, M.J. 2005. Genome-wide association studies for common diseases and complex traits. *Nature reviews genetics,* 6(2):95-108.

Ho, A.J., Stein, J.L., Hua, X., Lee, S., Hibar, D.P., Leow, A.D., Dinov, I.D., Toga, A.W., Saykin, A.J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M.J., Craig, D.W., Gerber, J.D., Allen, A.N., Corneveaux, J.J., Stephan, D.A., DeCarli, C.S., DeChairo, B.M., Potkin, S.G., Jack, C.R.,Jr, Weiner, M.W., Raji, C.A., Lopez, O.L., Becker, J.T.,

Carmichael, O.T., Thompson, P.M. & Alzheimer's Disease Neuroimaging Initiative. 2010. A commonly carried allele of the obesity-related FTO gene is associated with reduced brain volume in the healthy elderly. *Proceedings of the national academy of sciences of the United States of America,* 107(18):8404-8409.

Hochberg, Z., Feil, R., Constancia, M., Fraga, M., Junien, C., Carel, J., Boileau, P., Le Bouc, Y., Deal, C. & Lillycrop, K. 2010. Child health, developmental plasticity, and epigenetic programming. *Endocrine reviews,* 32(2):159-224.

Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., Riley, J., Purvis, I. & Xu, C.F. 2004. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *European journal of human genetics: EJHG,* 12(5):395-399.

Huang, J., Ellinghaus, D., Franke, A., Howie, B. & Li, Y. 2012. 1000 genomes-based imputation identifies novel and refined associations for the Wellcome Trust case control consortium phase 1 data. *European journal of human genetics,* 20(7):801-805.

Hugo, E.R., Brandebourg, T.D., Woo, J.G., Loftus, J., Alexander, J.W. & Ben-Jonathan, N. 2008. Bisphenol A at environmentally relevant doses inhibits adiponectin release from human adipose tissue explants and adipocytes. *Environmental health perspectives,* 116(12):1642-1647.

Humphrey, A., McCarty, D., Mackay, I., Rowley, M., Dwyer, T. & Zimmet, P. 1998. Autoantibodies to glutamic acid decarboxylase and phenotypic features associated with early insulin treatment in individuals with adult-onset diabetes mellitus. *Diabetic medicine,* 15(2):113-119.

International Expert Committee. 2009. International expert committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes care,* 32(7):1327-1334.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., Bras, J.M., Schymick, J.C., Hernandez, D.G., Traynor, B.J., Simon-Sanchez, J., Matarin, M., Britton, A., van de

Leemput, J., Rafferty, I., Bucan, M., Cann, H.M., Hardy, J.A., Rosenberg, N.A. & Singleton, A.B. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature,* 451(7181):998-1003.

Jarrett, R. J., Al Sayegh, H. 1978. Impaired glucose tolerance: Defining those at risk of diabetic complications. *Diabetologia,* 15243.

Jarrett, R.J. & Keen, H. 1976. Hyperglycaemia and diabetes mellitus. *Lancet,* 2(7993):1009-1012.

Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W.S. & Dixon, M.E. 1999. Deep common ancestry of Indian and western-eurasian mitochondrial DNA lineages. *Current biology,* 9(22):1331-1334.

Köbberling, J. & Tillil, H. 1982. Empirical risk figures for first degree relatives of non-insulin dependent diabetics. *The genetics of diabetes mellitus,* 201-209.

Koenig, R.J., Peterson, C.M., Jones, R.L., Saudek, C., Lehrman, M. & Cerami, A. 1976. Correlation of glucose regulation and hemoglobin A(ic) in diabetes mellitus. *New England journal of medicine,* 295(8):417-420.

Kommoju, U.J. & Reddy, B.M. 2011; 2011. Genetic etiology of type 2 diabetes mellitus: A review, *International journal of diabetes in developing countries,* 31(2):51 - 64.

Kooner, J.S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., Been, L.F., Chia, K.S., Dimas, A.S., Hassanali, N., Jafar, T., Jowett, J.B., Li, X., Radha, V., Rees, S.D., Takeuchi, F., Young, R., Aung, T., Basit, A., Chidambaram, M., Das, D., Grundberg, E., Hedman, A.K., Hydrie, Z.I., Islam, M., Khor, C.C., Kowlessur, S., Kristensen, M.M., Liju, S., Lim, W.Y., Matthews, D.R., Liu, J., Morris, A.P., Nica, A.C., Pinidiyapathirage, J.M., Prokopenko, I., Rasheed, A., Samuel, M., Shah, N., Shera, A.S., Small, K.S., Suo, C., Wickremasinghe, A.R., Wong, T.Y., Yang, M., Zhang, F., DIAGRAM, MuTHER, Abecasis, G.R., Barnett, A.H., Caulfield, M., Deloukas, P., Frayling, T.M., Froguel, P., Kato, N., Katulanda, P., Kelly, M.A., Liang, J., Mohan, V.,

Sanghera, D.K., Scott, J., Seielstad, M., Zimmet, P.Z., Elliott, P., Teo, Y.Y., McCarthy, M.I., Danesh, J., Tai, E.S. & Chambers, J.C. 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nature genetics,* 43(10):984-989.

Kotzé, H.J. 2010. Evaluation of specific genotypes in the context of the type 2 diabetes risk phenotype in the black South African population. Potchefstroom: NWU. (Thesis-PhD).

Larsen, N., Vogensen, F.K., Van Den Berg, F.W.J., Nielsen, D.S., Andreasen, A.S., Pedersen, B.K., Al-Soud, W.A., Sørensen, S.J., Hansen, L.H. & Jakobsen, M. 2010. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *Plos one,* 5(2).

Lettre, G., Lange, C. & Hirschhorn, J.N. 2007. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genetic epidemiology,* 31(4):358-362.

Levitt, N.S., Steyn, K., Lambert, E.V., Reagon, G., Lombard, C.J., Fourie, J.M., Rossouw, K. & Hoffman, M. 1999. Modifiable risk factors for type 2 diabetes mellitus in a peri-urban community in South Africa *Diabetic medicine: A journal of the British diabetic association,* 16(11):946-950.

Lewis, C.M. & Knight, J. 2012. Introduction to genetic association studies. *Cold spring harbor protocols,* 2012(3):297-306.

Li, M., Li, C. & Guan, W. 2008. Evaluation of coverage variation of SNP chips for genome-wide association studies. *European journal of human genetics: EJHG,* 16(5):635-643.

Lo, S.S.S., Tun, R.Y.M., Hawa, M. & Leslie, R.D.G. 1991. Studies of diabetic twins *Diabetes/metabolism reviews,* 7(4):223-238.

Lohman, T.G., Riche, A.F & Martorell, R. 1988. Anthropometric standardization reference manual.Champaighn, Ilinois: Human kinetic books, 177.

Lunetta, K.L. 2008. Genetic association studies. *Circulation,* 118(1):96-101.

Lupski, J.R. 2007. Structural variation in the human genome. *The New England journal of medicine,* 356(11):1169-1171.

Lyssenko, V., Almgren, P., Anevski, D., Perfekt, R., Lahti, K., Nissen, M., Isomaa, B., Forsen, B., Homstrom, N., Saloranta, C., Taskinen, M.R., Groop, L., Tuomi, T. & Botnia study group. 2005. Predictors of and longitudinal changes in insulin sensitivity and secretion preceding onset of type 2 diabetes. *Diabetes,* 54(1):166-174.

MacIntyre, U.E., Venter, C.S & Vorster, H.H 2001a. A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. relative validation by 7-day weighed records and biomarkers. *Public health nutrition,* 4(01):63-71.

MacIntyre, U.E., Venter, C.E., Vorster, H.H & Steyn, H.S. 2001b. A combination of statistical methods for the analysis of the relative validation data of the quantitative food frequency questionnaire used in the THUSA study. *Public health nutrition,* 4(01):45-51.

May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., Hartmann, N., Staedtler, F. & Ramsay, M. 2013. Genetic diversity in black South Africans from Soweto. *BMC genomics,* 14644-64

Mayer-Davis, E.J., Dabelea, D., Lamichhane, A.P., D'Agostino Jr., R.B., Liese, A.D., Thomas, J., Mckeown, R.E. & Hamman, R.F. 2008. Breast-feeding and type 2 diabetes in the youth of three ethnic groups: The SEARCH for diabetes in youth case-control study. *Diabetes care,* 31(3):470-475.

Mbanya, J.C., Motala, A.A., Sobngwi, E., Assah, F.K. & Enoru, S.T. 2010. Diabetes in sub-Saharan Africa *Lancet,* 375(9733):2254-2266.

McCance, D.R., Hanson, R.L., Charles, M., Jacobsson, L.T.H., Pettitt, D.J., Bennett, P.H. & Knowler, W.C. 1994. Comparison of tests for glycated haemoglobin and fasting and two hour plasma glucose concentrations as diagnostic methods for diabetes. *British medical journal,* 308(6940):1323-1328.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P. & Hirschhorn, J.N. 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature reviews genetics,* 9(5):356-369.

McCarthy, M.I. & Hirschhorn, J.N. 2008. Genome-wide association studies: Potential next steps on a genetic journey *Human molecular genetics,* 17(R2):R156-R165.

Medici, F., Hawa, M., Lanari, A., Pyke, D.A. & Leslie, R.D. 1999. Concordance rate for type II diabetes mellitus in monozygotic twins: Actuarial analysis. *Diabetologia,* 42(2):146-150.

Mesarovic, M.D., Sreenath, S.N. & Keene, J.D. 2004. Search for organising principles: Understanding in systems biology. *Systems biology,* 1(1):19-27.

Moore DS, M.G. 2003. Introduction to the practice of statistics. New York: W.H. Freeman and Company.

Müller, T.D., Tschöp, M.H. & Hofmann, S. 2013. Emerging function of fat mass and obesity-associated protein (FTO). *PLoS genetics,* 9(1):e1003223.

Nair, R.P., Duffin, K.C., Helms, C., Ding, J., Stuart, P.E., Goldgar, D., Gudjonsson, J.E., Li, Y., Tejasvi, T., Feng, B.-., Ruether, A., Schreiber, S., Weichenthal, M., Gladman, D., Rahman, P., Schrodi, S.J., Prahalad, S., Guthery, S.L., Fischer, J., Liao, W., Kwok, P.-., Menter, A., Lathrop, G.M., Wise, C.A., Begovich, A.B., Voorhees, J.J., Elder, J.T., Krueger, G.G., Bowcock, A.M. & Abecasis, G.R. 2009. Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways. *Nature genetics,* 41(2):199-204.

National Diabetes Data Group. 1979. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes,* 281038–1057.

Newman, B., Selby, J.V., King, M.C., Slemenda, C., Fabsitz, R. & Friedman, G.D. 1987. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia,* 30(10):763-768.

Nguyen, T.T., Wang, J.J. & Wong, T.Y. 2007. Retinal vascular changes in pre-diabetes and prehypertension: New findings and their research and clinical implications *Diabetes care,* 30(10):2708-2715.

Palmer, N.D., McDonough, C.W., Hicks, P.J., Roh, B.H., Wing, M.R., An, S.S., Hester, J.M., Cooke, J.N., Bostrom, M.A., Rudock, M.E., Talbert, M.E., Lewis, J.P., DIAGRAM Consortium, MAGIC Investigators, Ferrara, A., Lu, L., Ziegler, J.T., Sale, M.M., Divers, J., Shriner, D., Adeyemo, A., Rotimi, C.N., Ng, M.C., Langefeld, C.D., Freedman, B.I., Bowden, D.W., Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., McCulloch, L.J., Ferreira, T., Grallert, H., Amin, N., Wu, G., Willer, C.J., Raychaudhuri, S., McCarroll, S.A., Langenberg, C., Hofmann, O.M., Dupuis, J., Qi, L., Segre, A.V., van Hoek, M., Navarro, P., Ardlie, K., Balkau, B., Benediktsson, R., Bennett, A.J., Blagieva, R., Boerwinkle, E., Bonnycastle, L.L., Bostrom, K.B., Bravenboer, B., Bumpstead, S., Burtt, N.P., Charpentier, G., Chines, P.S., Cornelis, M., Couper, D.J., Crawford, G., Doney, A.S., Elliott, K.S., Elliott, A.L., Erdos, M.R., Fox, C.S., Franklin, C.S., Ganser, M., Gieger, C., Grarup, N., Green, T., Griffin, S., Groves, C.J., Guiducci, C., Hadjadj, S., Hassanali, N., Herder, C., Isomaa, B., Jackson, A.U., Johnson, P.R., Jorgensen, T., Kao, W.H., Klopp, N., Kong, A., Kraft, P., Kuusisto, J., Lauritzen, T., Li, M., Lieverse, A., Lindgren, C.M., Lyssenko, V., Marre, M., Meitinger, T., Midthjell, K., Morken, M.A., Narisu, N., Nilsson, P., Owen, K.R., Payne, F., Perry, J.R., Petersen, A.K., Platou, C., Proenca, C., Prokopenko, I., Rathmann, W., Rayner, N.W., Robertson, N.R., Rocheleau, G., Roden, M., Sampson, M.J., Saxena, R., Shields, B.M., Shrader, P., Sigurdsson, G., Sparso, T., Strassburger, K., Stringham, H.M., Sun, Q., Swift, A.J., Thorand, B., Tichet, J., Tuomi, T., van Dam, R.M., van Haeften, T.W., van Herpt, T., van Vliet-Ostaptchouk, J.V., Walters, G.B., Weedon, M.N., Wijmenga, C., Witteman, J.,

Bergman, R.N., Cauchi, S., Collins, F.S., Gloyn, A.L., Gyllensten, U., Hansen, T., Hide, W.A., Hitman, G.A., Hofman, A., Hunter, D.J., Hveem, K., Laakso, M., Mohlke, K.L., Morris, A.D., Palmer, C.N., Pramstaller, P.P., Rudan, I., Sijbrands, E., Stein, L.D., Tuomilehto, J., Uitterlinden, A., Walker, M., Wareham, N.J., Watanabe, R.M., Abecasis, G.R., Boehm, B.O., Campbell, H., Daly, M.J., Hattersley, A.T., Hu, F.B., Meigs, J.B., Pankow, J.S., Pedersen, O., Wichmann, H.E., Barroso, I., Florez, J.C., Frayling, T.M., Groop, L., Sladek, R., Thorsteinsdottir, U., Wilson, J.F., Illig, T., Froguel, P., van Duijn, C.M., Stefansson, K., Altshuler, D., Boehnke, M., McCarthy, M.I., Soranzo, N., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Magi, R., Randall, J., Johnson, T., Elliott, P., Rybin, D., Henneman, P., Dehghan, A., Hottenga, J.J., Song, K., Goel, A., Egan, J.M., Lajunen, T., Doney, A., Kanoni, S., Cavalcanti-Proenca, C., Kumari, M., Timpson, N.J., Zabena, C., Ingelsson, E., An, P., O'Connell, J., Luan, J., Elliott, A., McCarroll, S.A., Roccasecca, R.M., Pattou, F., Sethupathy, P., Ariyurek, Y., Barter, P., Beilby, J.P., Ben-Shlomo, Y., Bergmann, S., Bochud, M., Bonnefond, A., Borch-Johnsen, K., Bottcher, Y., Brunner, E., Bumpstead, S.J., Chen, Y.D., Chines, P., Clarke, R., Coin, L.J., Cooper, M.N., Crisponi, L., Day, I.N., de Geus, E.J., Delplanque, J., Fedson, A.C., Fischer-Rosinsky, A., Forouhi, N.G., Frants, R., Franzosi, M.G., Galan, P., Goodarzi, M.O., Graessler, J., Grundy, S., Gwilliam, R., Hallmans, G., Hammond, N., Han, X., Hartikainen, A.L., Hayward, C., Heath, S.C., Hercberg, S., Hicks, A.A., Hillman, D.R., Hingorani, A.D., Hui, J., Hung, J., Jula, A., Kaakinen, M., Kaprio, J., Kesaniemi, Y.A., Kivimaki, M., Knight, B., Koskinen, S., Kovacs, P., Kyvik, K.O., Lathrop, G.M., Lawlor, D.A., Le Bacquer, O., Lecoeur, C., Li, Y., Mahley, R., Mangino, M., Manning, A.K., Martinez-Larrad, M.T., McAteer, J.B., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B.D., Mukherjee, S., Naitza, S., Neville, M.J., Oostra, B.A., Orru, M., Pakyz, R., Paolisso, G., Pattaro, C., Pearson, D., Peden, J.F., Pedersen, N.L., Perola, M., Pfeiffer, A.F., Pichler, I., Polasek, O., Posthuma, D., Potter, S.C., Pouta, A., Province, M.A., Psaty, B.M., Rayner, N.W., Rice, K., Ripatti, S., Rivadeneira, F., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A.A., Scheet, P., Seedorf, U., Sharp, S.J., Shields, B., Sijbrands, E.J., Silveira, A., Simpson, L., Singleton, A., Smith, N.L., Sovio, U., Swift, A., Syddall, H., Syvanen, A.C., Tanaka, T., Tonjes, A., Uitterlinden, A.G., van Dijk, K.W., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P.J., Walley, A., Ward, K.L., Watkins, H., Wild, S.H., Willemsen, G., Witteman, J.C., Yarnell, J.W., Zelenika, D., Zethelius, B., Zhai, G., Zhao, J.H., Zillikens, M.C.,

Borecki, I.B., Loos, R.J., Meneton, P., Magnusson, P.K., Nathan, D.M., Williams, G.H., Silander, K., Salomaa, V., Smith, G.D., Bornstein, S.R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A.R., Cooper, C., Dedoussis, G.V., Serrano-Rios, M., Lind, L., Palmer, L.J., Franks, P.W., Ebrahim, S., Marmot, M., Kao, W.H., Pramstaller, P.P., Wright, A.F., Stumvoll, M., Hamsten, A., Buchanan, T.A., Valle, T.T., Rotter, J.I., Siscovick, D.S., Penninx, B.W., Boomsma, D.I., Deloukas, P., Spector, T.D., Ferrucci, L., Cao, A., Scuteri, A., Schlessinger, D., Uda, M., Ruokonen, A., Jarvelin, M.R., Waterworth, D.M., Vollenweider, P., Peltonen, L., Mooser, V. & Sladek, R. 2012. A genome-wide association search for type 2 diabetes genes in African Americans. *PloS one,* 7(1):e29202.

Parra, E., Below, J., Krithika, S., Valladares, A., Barta, J., Cox, N., Hanis, C., Wacher, N., Garcia-Mena, J. & Hu, P. 2011. Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. *Diabetologia,* 54(8):2038-2046.

Peer, N., Steyn, K., Lombard, C., Lambert, E.V., Vythilingum, B. & Levitt, N.S. 2012. Rising diabetes prevalence among urban-dwelling black South Africans *PloS one,* 7(9):e43336.

Petretto, E., Liu, E.T. & Aitman, T.J. 2007. A gene harvest revealing the archeology and complexity of human disease. *Nature genetics,* 39(11):1299-1301.

Pettitt, D.J., Lawrence, J.M., Beyer, J., Hillier, T.A., Liese, A.D., Mayer-Davis, B., Loots, B., Imperatore, G., Liu, L., Dolan, L.M., Linder, B. & Dabelea, D. 2008. Association between maternal diabetes in utero and age at offspring's diagnosis of type 2 diabetes. *Diabetes care,* 31(11):2126-2130.

Phipps, K., Barker, D.J.P., Hales, C.N., Fall, C.H.D., Osmond, C. & Clark, P.M.S. 1993. Fetal growth and impaired glucose tolerance in men and women. *Diabetologia,* 36(3):225-228.

Pirie, F., Motala, A., Pegoraro, R., Paruk, I., Govender, T. & Rom, L. 2010. Variants in PPARG, KCNJ11, TCF7L2, FTO, and HHEX genes in South African subjects of Zulu descent with type 2 diabetes. *African journal of diabetes medicine,* 18(1):12-16.

Purcell, S., Cherny, S.S. & Sham, P.C. 2003. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics (Oxford, England),* 19(1):149-150.

Purcell, S., Neale, B., Todd-brown, K., Thomas, l., Ferreira, M.A, Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. & Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics,* 81559-575.

Qi, L., Cornelis, M.C., Zhang, C., van Dam, R.M. & Hu, F.B. 2009. Genetic predisposition, western dietary pattern, and the risk of type 2 diabetes in men. *The American journal of clinical nutrition,* 89(5):1453-1458.

Qi, Q. & Hu, F.B. 2012. Genetics of type 2 diabetes in European populations. *Journal of diabetes,* 4(3):203-212.

Rakyan, V. & Whitelaw, E. 2003. Transgenerational epigenetic inheritance. *Current biology: CB,* 13(1):.

Ramsay, M., Tiemessen, C.T., Choudhury, A. & Soodyall, H. 2011. Africa: The next frontier for human disease gene discovery? *Human molecular genetics,* 20(R2):R214-20.

Ramsay, M. 2012. Africa: Continent of genome contrasts with implications for biomedical research and health. *FEBS letters,* 586(18):2813-2819.

Reaven, G.M., Hollenbeck, C.B. & Chen, Y.-I. 1989. Relationship between glucose tolerance, insulin secretion, and insulin action in non-obese individuals with varying degrees of glucose tolerance. *Diabetologia,* 32(1):52-55.

Reich, D.E. & Lander, E.S. 2001. On the allelic spectrum of human disease. *Trends in genetics : TIG,* 17(9):502-510.

Remillard, R.B.J. & Bunce, N.J. 2002. Linking dioxins to diabetes: Epidemiology and biologic plausibility. *Environmental health perspectives,* 110(9):853-858.

Reppert, S.M., Godson, C., Mahle, C.D., Weaver, D.R., Slaugenhaupt, S.A. & Gusella, J.F. 1995. Molecular characterization of a second melatonin receptor expressed in human retina and brain: The Mel1b melatonin receptor. *Proceedings of the national academy of sciences of the United States of America,* 92(19):8734-8738.

Risch, N. 1990. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *American journal of human genetics,* 46(2):242-253.

Rosenberg, N.A., Huang, L., Jewett, E.M., Szpiech, Z.A., Jankovic, I. & Boehnke, M. 2010. Genome-wide association studies in diverse populations. *Nature reviews genetics,* 11(5):356-366.

Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., International Inflammatory Bowel Disease Genetics Constortium, Cotsapas, C. & Daly, M.J. 2011. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS genetics,* 7(1):e1001273.

Rotimi, C.N., Chen, G., Adeyemo, A.A., Furbert-Harris, P., Parish-Gause, D., Zhou, J., Berg, K., Adegoke, O., Amoah, A., Owusu, S., Acheampong, J., Agyenim-Boateng, K., Eghan, B.A.,Jr, Oli, J., Okafor, G., Ofoegbu, E., Osotimehin, B., Abbiyesuku, F., Johnson, T., Rufus, T., Fasanmade, O., Kittles, R., Daniel, H., Chen, Y., Dunston, G., Collins, F.S. & Africa America Diabetes Mellitus (AADM) Study. 2004. A genome-wide search for type 2 diabetes susceptibility genes in West Africans: The Africa America diabetes mellitus (AADM) study. *Diabetes,* 53(3):838-841.

Rung, J., Cauchi, S., Albrechtsen, A., Shen, L., Rocheleau, G., Cavalcanti-Proença, C., Bacot, F., Balkau, B., Belisle, A., Borch-Johnsen, K., Charpentier, G., Dina, C., Durand, E., Elliott, P., Hadjadj, S., Järvelin, M.-., Laitinen, J., Lauritzen, T., Marre, M., Mazur, A., Meyre, D., Montpetit, A., Pisinger, C., Posner, B., Poulsen, P., Pouta, A., Prentki, M., Ribel-Madsen, R., Ruokonen, A., Sandbaek, A., Serre, D., Tichet, J., Vaxillaire, M., Wojtaszewski, J.F.P., Vaag, A., Hansen, T., Polychronakos, C., Pedersen, O., Froguel, P. & Sladek, R. 2009. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nature genetics,* 41(10):1110-1115.

Saad, M.F., Knowler, W.C., Pettitt, D.J., Nelson, R.G. & Bennett, P.H. 1988. Transient impaired glucose tolerance in Pima Indians: Is it important? *British medical journal,* 297(6661):1438-1441.

Sacks, D.B., Burns, D.E., Goldstein, D.E., Maclaren, N.K., McDonald, J.M. & Parrott, M. 2002. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Clinical chemistry,* 48(3):436-472.

Sandu, O., Song, K., Cai, W., Zheng, F., Uribarri, J. & Vlassara, H. 2005. Insulin resistance and type 2 diabetes in high-fat-fed mice are linked to high glycotoxin intake. *Diabetes,* 54(8):2314-2319.

Sanger, F., Nicklen, S. & Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences,* 74(12):5463-5467.

Sanghera, D.K. & Blackett, P.R. 2012. Type 2 diabetes genetics: Beyond GWAS *journal of diabetes & metabolism,* 3(198):6948.

Saudek, C.D., Herman, W.H., Sacks, D.B., Bergenstal, R.M., Edelman, D. & Davidson, M.B. 2008. A new look at screening and diagnosing diabetes mellitus. *The journal of clinical endocrinology and metabolism,* 93(7):2447-2453.

Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. & Snyder, M. 2012. Linking disease associations with regulatory information in the human genome. *Genome research,* 22(9):1748-1759.

Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., Harris, R.S., Petersen, D.C., Zhao, F. & Qi, J. 2010. Complete Khoisan and Bantu genomes from Southern Africa. *Nature,* 463(7283):943-947.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U., Prokunina-Olsson, L., Ding, C.-., Swift, A.J., Narisu, Hu, T., Pruim, R., Xiao, R., Li, X.-., Conneely, K.N., Riebow, N.L., Sprau, A.G., Tong, M., White, P.P., Hetrick, K.N., Barnhart, M.W., Bark, C.W., Goldstein, J.L., Watkins, L., Xiang, F., Saramies, J., Buchanan, T.A., Watanabe, R.M., Valle, T.T., Kinnunen, L., Abecasis, G.R., Pugh, E.W., Doheny, K.F., Bergman, R.N., Tuomilehto, J., Collins, F.S. & Boehnke, M. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science,* 316(5829):1341-1345.

Sesti, G. 2006. Pathophysiology of insulin resistance. *Best practice & research clinical endocrinology & metabolism,* 20(4):665-679.

Shu, X.O., Long, J., Cai, Q., Qi, L., Xiang, Y.B., Cho, Y.S., Tai, E.S., Li, X., Lin, X., Chow, W.H., Go, M.J., Seielstad, M., Bao, W., Li, H., Cornelis, M.C., Yu, K., Wen, W., Shi, J., Han, B.G., Sim, X.L., Liu, L., Qi, Q., Kim, H.L., Ng, D.P., Lee, J.Y., Kim, Y.J., Li, C., Gao, Y.T., Zheng, W. & Hu, F.B. 2010. Identification of new genetic risk variants for type 2 diabetes. *PLoS genetics,* 6(9):e1001127.

Sim, X., Ong, R.T., Suo, C., Tay, W.T., Liu, J., Ng, D.P., Boehnke, M., Chia, K.S., Wong, T.Y., Seielstad, M., Teo, Y.Y. & Tai, E.S. 2011. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS genetics,* 7(4):e1001363.

Skinner, M.K., Manikkam, M. & Guerrero-Bosagna, C. 2010. Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends in endocrinology and metabolism,* 21(4):214-222.

Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., Balkau, B., Heude, B., Charpentier, G., Hudson, T.J., Montpetit, A., Pshezhetsky, A.V., Prentki, M., Posner, B.I., Balding, D.J., Meyre, D., Polychronakos, C. & Froguel, P. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature,* 445(7130):881-885.

Smemo, S., Tena, J.J., Kim, K., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R. & Wasserman, N.F. 2014. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature.*

Sorkin, J.D., Muller, D.C., Fleg, J.L. & Andres, R. 2005. The relation of fasting and 2-h postchallenge plasma glucose concentrations to mortality: Data from the Baltimore longitudinal study of aging with a critical review of the literature. *Diabetes care,* 28(11):2626-2632.

Suhre, K., Shin, S.Y., Petersen, A.K., Mohney, R.P., Meredith, D., Wagele, B., Altmaier, E., CARDIoGRAM, Deloukas, P., Erdmann, J., Grundberg, E., Hammond, C.J., de Angelis, M.H., Kastenmuller, G., Kottgen, A., Kronenberg, F., Mangino, M., Meisinger, C., Meitinger, T., Mewes, H.W., Milburn, M.V., Prehn, C., Raffler, J., Ried, J.S., Romisch-Margl, W., Samani, N.J., Small, K.S., Wichmann, H.E., Zhai, G., Illig, T., Spector, T.D., Adamski, J., Soranzo, N. & Gieger, C. 2011. Human metabolic individuality in biomedical and pharmaceutical research. *Nature,* 477(7362):54-60.

Sumner, C.J., Sheth, S., Griffin, J.W., Cornblath, D.R. & Polydefkis, M. 2003. The spectrum of neuropathy in diabetes and impaired glucose tolerance *Neurology,* 60(1):108-111.

Taneera, J., Lang, S., Sharma, A., Fadista, J., Zhou, Y., Ahlqvist, E., Jonsson, A., Lyssenko, V., Vikman, P., Hansson, O., Parikh, H., Korsgren, O., Soni, A., Krus, U., Zhang, E., Jing, X., Esguerra, J.L.S., Wollheim, C.B., Salehi, A., Rosengren, A., Renström, E. & Groop, L. 2012. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell metabolism,* 16(1):122-134.

Teo, Y.Y., Small, K.S. & Kwiatkowski, D.P. 2010. Methodological challenges of genome-wide association analysis in Africa. *Nature reviewsgenetics,* 11(2):149-160.

Tong, Y., Lin, Y., Zhang, Y., Yang, J., Zhang, Y., Liu, H. & Zhang, B. 2009. Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: A large human genome epidemiology (HuGE) review and meta-analysis. *BMC medical genetics,* 10(1):15.

Tsai, F.J., Yang, C.F., Chen, C.C., Chuang, L.M., Lu, C.H., Chang, C.T., Wang, T.Y., Chen, R.H., Shiu, C.F., Liu, Y.M., Chang, C.C., Chen, P., Chen, C.H., Fann, C.S., Chen, Y.T. & Wu, J.Y. 2010. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS genetics,* 6(2):e1000847.

Tsuji, S. 2010. Genetics of neurodegenerative diseases: Insights from high-throughput resequencing. *Human molecular genetics,* 19(R1):R65-70.

Turek, F.W., Joshu, C., Kohsaka, A., Lin, E., Ivanova, G., McDearmon, E., Laposky, A., Losee-Olson, S., Easton, A., Jensen, D.R., Eckel, R.H., Takahashi, J.S. & Bass, J. 2005. Obesity and metabolic syndrome in circadian clock mutant mice. *Science (New York, N.Y.),* 308(5724):1043-1045.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. & Gordon, J.I. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature,* 444(7122):1027-1031.

Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P., Holmkvist, J., Borch-Johnsen, K., Jorgensen, T., Sandbaek, A., Lauritzen, T., Hansen, T., Nurbaya, S., Tsunoda, T., Kubo, M., Babazono, T., Hirose, H., Hayashi, M., Iwamoto, Y., Kashiwagi, A., Kaku, K., Kawamori, R., Tai, E.S., Pedersen, O., Kamatani, N., Kadowaki, T., Kikkawa, R., Nakamura, Y. & Maeda, S. 2008. SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nature genetics,* 40(9):1098-1102.

Vassy, J.L. & Meigs, J.B. 2012. Is genetic testing useful to predict type 2 diabetes? *Best practice & research, clinical endocrinology & metabolism,* 26(2):189-201.

Verge, C.F., Gianani, R., Kawasaki, E., Yu, L., Pietropaolo, M., Jackson, R.A., Chase, H.P. & Eisenbarth, G.S. 1996. Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. *Diabetes,* 45(3 SUPPL.):926-933.

Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., McCulloch, L.J., Ferreira, T., Grallert, H., Amin, N., Wu, G., Willer, C.J., Raychaudhuri, S., McCarroll, S.A., Langenberg, C., Hofmann, O.M., Dupuis, J., Qi, L., Segrè, A.V., Van Hoek, M., Navarro, P., Ardlie, K., Balkau, B., Benediktsson, R., Bennett, A.J., Blagieva, R., Boerwinkle, E., Bonnycastle, L.L., Boström, K.B., Bravenboer, B., Bumpstead, S., Burtt, N.P., Charpentier, G., Chines, P.S., Cornelis, M., Couper, D.J., Crawford, G., Doney, A.S.F., Elliott, K.S., Elliott, A.L., Erdos, M.R., Fox, C.S., Franklin, C.S., Ganser, M., Gieger, C., Grarup, N., Green, T., Griffin, S., Groves, C.J., Guiducci, C., Hadjadj, S., Hassanali, N., Herder, C., Isomaa, B., Jackson, A.U., Johnson, P.R.V., Jørgensen, T., Kao, W.H.L., Klopp, N., Kong, A., Kraft, P., Kuusisto, J., Lauritzen, T., Li, M., Lieverse, A., Lindgren, C.M., Lyssenko, V., Marre, M., Meitinger, T., Midthjell, K., Morken, M.A., Narisu, N., Nilsson, P., Owen, K.R., Payne, F., Perry, J.R.B., Petersen, A.-., Platou, C., Proença, C., Prokopenko, I., Rathmann, W., Rayner, N.W., Robertson, N.R., Rocheleau, G., Roden, M., Sampson, M.J., Saxena, R., Shields, B.M., Shrader, P., Sigurdsson, G., Sparsø, T., Strassburger, K., Stringham, H.M., Sun, Q., Swift, A.J., Thorand, B., Tichet, J., Tuomi, T., Van Dam, R.M., Van Haeften, T.W., Van Herpt, T., Van Vliet-Ostaptchouk, J.V., Walters, G.B., Weedon, M.N., Wijmenga, C., Witteman, J., Bergman, R.N., Cauchi, S., Collins, F.S., Gloyn, A.L., Gyllensten, U., Hansen, T., Hide, W.A., Hitman, G.A., Hofman, A., Hunter, D.J., Hveem, K., Laakso, M., Mohlke, K.L., Morris, A.D., Palmer, C.N.A., Pramstaller, P.P., Rudan, I., Sijbrands, E., Stein, L.D., Tuomilehto, J., Uitterlinden, A., Walker, M., Wareham, N.J., Watanabe, R.M., Abecasis, G.R., Boehm, B.O., Campbell, H., Daly, M.J., Hattersley, A.T., Hu, F.B., Meigs, J.B., Pankow, J.S., Pedersen, O., Wichmann, H.-., Barroso, I., Florez, J.C., Frayling, T.M., Groop, L., Sladek, R., Thorsteinsdottir, U., Wilson, J.F., Illig, T., Froguel, P., Van Duijn, C.M.,

Stefansson, K., Altshuler, D., Boehnke, M. & McCarthy, M.I. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature genetics,* 42(7):579-589.

Warton, K., Foster, N.C., Gold, W.A. & Stanley, K.K. 2004. A novel gene family induced by acute inflammation in endothelial cells. *Gene,* 342(1):85-95.

Waters, K.M., Stram, D.O., Hassanein, M.T., Le Marchand, L., Wilkens, L.R., Maskarinec, G., Monroe, K.R., Kolonel, L.N., Altshuler, D., Henderson, B.E. & Haiman, C.A. 2010. Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups *PLoS genetics,* 6(8):10.1371/journal.pgen.1001078.

Wellcome Trust Case Control Consortium, Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E., Holmes, C., Marchini, J.L., Stirrups, K., Tobin, M.D., Wain, L.V., Yau, C., Aerts, J., Ahmad, T., Andrews, T.D., Arbury, H., Attwood, A., Auton, A., Ball, S.G., Balmforth, A.J., Barrett, J.C., Barroso, I., Barton, A., Bennett, A.J., Bhaskar, S., Blaszczyk, K., Bowes, J., Brand, O.J., Braund, P.S., Bredin, F., Breen, G., Brown, M.J., Bruce, I.N., Bull, J., Burren, O.S., Burton, J., Byrnes, J., Caesar, S., Clee, C.M., Coffey, A.J., Connell, J.M., Cooper, J.D., Dominiczak, A.F., Downes, K., Drummond, H.E., Dudakia, D., Dunham, A., Ebbs, B., Eccles, D., Edkins, S., Edwards, C., Elliot, A., Emery, P., Evans, D.M., Evans, G., Eyre, S., Farmer, A., Ferrier, I.N., Feuk, L., Fitzgerald, T., Flynn, E., Forbes, A., Forty, L., Franklyn, J.A., Freathy, R.M., Gibbs, P., Gilbert, P., Gokumen, O., Gordon-Smith, K., Gray, E., Green, E., Groves, C.J., Grozeva, D., Gwilliam, R., Hall, A., Hammond, N., Hardy, M., Harrison, P., Hassanali, N., Hebaishi, H., Hines, S., Hinks, A., Hitman, G.A., Hocking, L., Howard, E., Howard, P., Howson, J.M., Hughes, D., Hunt, S., Isaacs, J.D., Jain, M., Jewell, D.P., Johnson, T., Jolley, J.D., Jones, I.R., Jones, L.A., Kirov, G., Langford, C.F., Lango-Allen, H., Lathrop, G.M., Lee, J., Lee, K.L., Lees, C., Lewis, K., Lindgren, C.M., Maisuria-Armer, M., Maller, J., Mansfield, J., Martin, P., Massey, D.C., McArdle, W.L., McGuffin, P., McLay, K.E., Mentzer, A., Mimmack, M.L., Morgan, A.E., Morris, A.P., Mowat, C., Myers, S., Newman, W., Nimmo, E.R., O'Donovan, M.C., Onipinla, A., Onyiah, I., Ovington, N.R., Owen, M.J., Palin, K., Parnell, K., Pernet, D., Perry, J.R., Phillips, A.,

Pinto, D., Prescott, N.J., Prokopenko, I., Quail, M.A., Rafelt, S., Rayner, N.W., Redon, R., Reid, D.M., Renwick, Ring, S.M., Robertson, N., Russell, E., St Clair, D., Sambrook, J.G., Sanderson, J.D., Schuilenburg, H., Scott, C.E., Scott, R., Seal, S., Shaw-Hawkins, S., Shields, B.M., Simmonds, M.J., Smyth, D.J., Somaskantharajah, E., Spanova, K., Steer, S., Stephens, J., Stevens, H.E., Stone, M.A., Su, Z., Symmons, D.P., Thompson, J.R., Thomson, W., Travers, M.E., Turnbull, C., Valsesia, A., Walker, M., Walker, N.M., Wallace, C., Warren-Perry, M., Watkins, N.A., Webster, J., Weedon, M.N., Wilson, A.G., Woodburn, M., Wordsworth, B.P., Young, A.H., Zeggini, E., Carter, N.P., Frayling, T.M., Lee, C., McVean, G., Munroe, P.B., Palotie, A., Sawcer, S.J., Scherer, S.W., Strachan, D.P., Tyler-Smith, C., Brown, M.A., Burton, P.R., Caulfield, M.J., Compston, A., Farrall, M., Gough, S.C., Hall, A.S., Hattersley, A.T., Hill, A.V., Mathew, C.G., Pembrey, M., Satsangi, J., Stratton, M.R., Worthington, J., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W., Parkes, M., Rahman, N., Todd, J.A., Samani, N.J. & Donnelly, P. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature,* 464(7289):713-720.

WHO Expert Committee on Diabetes Mellitus. 1980. WHO expert committee on diabetes mellitus 2nd report. Geneva: WHO. (no. WHO Technical Report Series 646).

Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. 2005. A note on exact tests of hardy-weinberg equilibrium. *The American journal of human genetics,* 76(5):887-893

Wing, R.R., Blair, E.H., Bononi, P., Marcus, M.D., Watanabe, R. & Bergman, R.N. 1994. Caloric restriction per se is a significant factor in improvements in glycemic control and insulin sensitivity during weight loss in obese NIDDM patients. *Diabetes care,* 17(1):30-36

Yajnik, C.S., Deshpande, S.S., Jackson, A.A., Refsum, H., Rao, S., Fisher, D.J., Bhat, D.S., Naik, S.S., Coyaji, K.J., Joglekar, C.V., Joshi, N., Lubree, H.G., Deshpande, V.U., Rege, S.S. & Fall, C.H.D. 2008. Vitamin B12 and folate concentrations during pregnancy and insulin resistance in the offspring: The pune maternal nutrition study. *Diabetologia,* 51(1):29-38.

Yamauchi, T., Hara, K., Maeda, S., Yasuda, K., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., Grarup, N. & Cauchi, S. 2010. A genome-wide association

study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nature genetics,* 42(10):864-868.

Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., Hirota, Y., Mori, H., Jonsson, A., Sato, Y., Yamagata, K., Hinokio, Y., Wang, H.Y., Tanahashi, T., Nakamura, N., Oka, Y., Iwasaki, N., Iwamoto, Y., Yamada, Y., Seino, Y., Maegawa, H., Kashiwagi, A., Takeda, J., Maeda, E., Shin, H.D., Cho, Y.M., Park, K.S., Lee, H.K., Ng, M.C., Ma, R.C., So, W.Y., Chan, J.C., Lyssenko, V., Tuomi, T., Nilsson, P., Groop, L., Kamatani, N., Sekine, A., Nakamura, Y., Yamamoto, K., Yoshida, T., Tokunaga, K., Itakura, M., Makino, H., Nanjo, K., Kadowaki, T. & Kasuga, M. 2008. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nature genetics,* 40(9):1092-1097.

Zeggini, E., Scott, L.J., Saxena, R. & Voight, B.F. 2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature genetics,* 40(5):638-645.

# APPENDIX A

## Additive genetic model univariate and multivariate results for the selected SNPs

This appendix contains the univariate and multivate test results for the additive genetic model that are illustrated in Table A1.

**Table A.1 Univariate test results for the additive genetic model**

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | EMP1 |
|---|---|---|---|---|
| 1 | rs7542900 | A | G | 0.27 |
| 1 | rs10923931 | A | C | 0.22 |
| 2 | rs7578597 | G | A | 0.27 |
| 2 | rs243021 | A | G | 1.00 |
| 2 | rs7560163 | C | G | 1.00 |
| 2 | rs3923113 | A | C | 1.00 |
| 2 | rs7578326 | G | A | 0.91 |
| 2 | rs2943641 | A | G | 0.67 |
| 3 | rs17036101 | A | G | 0.21 |
| 3 | rs358806 | A | C | 0.13 |
| 3 | rs831571 | A | G | 0.04 |
| 3 | rs4607103 | A | G | 0.39 |
| 3 | rs3773506 | G | C | 0.64 |
| 3 | rs7630877 | A | G | 0.85 |
| 3 | rs4402960 | C | A | 0.57 |
| 3 | rs16861329 | A | G | 0.80 |
| 4 | rs6815464 | G | C | 0.97 |
| 4 | rs4689388 | G | A | 0.84 |
| 5 | rs4457053 | G | A | 0.97 |
| 5 | rs12518099 | G | A | 0.23 |
| 6 | rs7754840 | G | C | 0.06 |
| 6 | rs9470794 | G | A | 0.78 |
| 6 | rs9472138 | A | G | 0.05 |
| 6 | rs1048886 | G | A | 0.92 |
| 7 | rs864745 | G | A | 0.66 |
| 7 | rs7636 | A | G | 0.97 |
| 7 | rs6467136 | A | G | 0.41 |

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | EMP1 |
|---|---|---|---|---|
| 7 | rs972283 | A | G | 0.17 |
| 8 | rs896854 | G | A | 0.56 |
| 8 | rs13266634 | A | G | 0.94 |
| 9 | rs7041847 | G | A | 0.89 |
| 9 | rs17584499 | A | G | 1.00 |
| 9 | rs564398 | G | A | 0.26 |
| 9 | rs7018475 | C | A | 0.75 |
| 9 | rs13292136 | A | G | 0.73 |
| 10 | rs12779790 | G | A | 0.32 |
| 10 | rs6583826 | G | A | 0.19 |
| 10 | rs1111875 | A | G | 0.80 |
| 10 | rs7903146 | A | G | 0.61 |
| 10 | rs10741243 | G | C | 0.68 |
| 11 | rs231362 | A | G | 0.44 |
| 11 | rs2722769 | C | G | 0.36 |
| 11 | rs1387153 | A | G | 0.08 |
| 11 | rs7107217 | C | A | 0.76 |
| 12 | rs1153188 | A | T | 0.30 |
| 12 | rs1531343 | G | C | 0.26 |
| 12 | rs7961581 | G | A | 0.89 |
| 12 | rs7957197 | A | T | 0.69 |
| 13 | rs1359790 | A | G | 0.60 |
| 14 | rs730570 | A | G | 0.62 |
| 15 | rs7172432 | A | G | 0.65 |
| 15 | rs1436955 | G | A | 0.05 |
| 15 | rs7178572 | A | G | 0.39 |
| 15 | rs11634397 | G | A | 0.57 |
| 15 | rs2028299 | C | A | 0.23 |
| 15 | rs8042680 | C | A | 0.48 |
| 16 | rs8050136 | A | C | 0.47 |
| 16 | rs17797882 | A | G | 0.07 |
| 16 | rs16955379 | A | G | 0.76 |
| 17 | rs391300 | G | A | 0.16 |
| 17 | rs4430796 | A | G | 0.30 |
| 18 | rs10460009 | A | G | 0.65 |
| 19 | rs472265 | G | A | 0.88 |
| 20 | rs4812829 | A | G | 0.50 |

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | EMP1 |
|---|---|---|---|---|
| 21 | rs2833610 | A | G | 0.66 |
| 23 | rs5945326 | G | A | 0.75 |

SNP=single nucleotide polymorphisms; A = adenine; G = guanine; C = cytosine; T = thymine; EMP1 = point-wise empirical value

# APPENDIX B

## General genotypic genetic model univariate results for the selected SNPs

This appendix contains the univariate test results for the general genotypic genetic model that are indicated in Table B.1.

**Table B.1 Univariate tests results of the general genotypic genetic model**

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 1 | rs7542900 | A | G | 24/86/53 | 42/69/57 | 0.03 |
| 1 | rs10923931 | A | C | 10/50/103 | 11/65/94 | 0.32 |
| 2 | rs7578597 | G | A | 8/67/88 | 7/60/102 | 0.52 |
| 2 | rs243021 | A | G | 27/94/42 | 34/87/49 | 0.48 |
| 2 | rs7560163 | C | G | 11/41/109 | 7/52/109 | 0.36 |
| 2 | rs3923113 | A | C | 5/39/119 | 3/45/122 | 0.68 |
| 2 | rs7578326 | G | A | 28/84/50 | 34/80/55 | 0.69 |
| 2 | rs2943641 | A | G | 19/66/64 | 14/72/65 | 0.61 |
| 3 | rs17036101 | A | G | 1/36/126 | 4/43/123 | 0.32 |
| 3 | rs358806 | A | C | 0/7/119 | 0/13/104 | 0.10 |
| 3 | rs831571 | A | G | 1/28/134 | 6/38/126 | 0.07 |
| 3 | rs4607103 | A | G | 19/72/72 | 26/74/70 | 0.62 |
| 3 | rs3773506 | G | C | 3/37/123 | 3/34/133 | 0.88 |
| 3 | rs7630877 | A | G | 20/73/48 | 18/74/48 | 0.96 |
| 3 | rs4402960 | C | A | 31/84/48 | 30/85/55 | 0.87 |
| 3 | rs16861329 | A | G | 0/6/157 | 0/8/162 | 0.78 |
| 4 | rs6815464 | G | C | 3/35/124 | 3/36/129 | 0.98 |
| 4 | rs4689388 | G | A | 11/76/76 | 9/81/79 | 0.86 |
| 5 | rs4457053 | G | A | 11/64/87 | 14/61/94 | 0.73 |
| 5 | rs12518099 | G | A | 7/72/83 | 12/78/76 | 0.40 |
| 6 | rs7754840 | G | C | 19/87/28 | 12/88/40 | 0.18 |
| 6 | rs9470794 | G | A | 13/58/92 | 13/65/92 | 0.89 |
| 6 | rs9472138 | A | G | 11/68/84 | 17/82/69 | 0.13 |
| 6 | rs1048886 | G | A | 11/75/77 | 14/72/84 | 0.77 |
| 7 | rs864745 | G | A | 4/45/114 | 4/42/123 | 0.89 |

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 7 | rs7636 | A | G | 20/63/80 | 21/64/84 | 0.99 |
| 7 | rs6467136 | A | G | 7/64/83 | 7/55/91 | 0.64 |
| 7 | rs972283 | A | G | 2/37/124 | 2/27/141 | 0.26 |
| 8 | rs896854 | G | A | 4/55/94 | 6/57/90 | 0.80 |
| 8 | rs13266634 | A | G | 4/24/126 | 9/15/138 | 0.13 |
| 9 | rs7041847 | G | A | 1/19/143 | 0/23/147 | 0.62 |
| 9 | rs17584499 | A | G | 1/22/140 | 0/24/138 | 0.81 |
| 9 | rs564398 | G | A | 0/5/158 | 0/2/167 | 0.29 |
| 9 | rs7018475 | C | A | 1/37/123 | 2/33/130 | 0.65 |
| 9 | rs13292136 | A | G | 2/12/149 | 0/14/156 | 0.50 |
| 10 | rs12779790 | G | A | 2/29/131 | 2/39/128 | 0.46 |
| 10 | rs6583826 | G | A | 28/86/42 | 38/89/37 | 0.43 |
| 10 | rs1111875 | A | G | 11/50/102 | 13/53/104 | 0.94 |
| 10 | rs7903146 | A | G | 15/61/85 | 12/77/79 | 0.31 |
| 10 | rs10741243 | G | C | 11/73/77 | 15/74/79 | 0.81 |
| 11 | rs231362 | A | G | 5/47/109 | 5/42/123 | 0.64 |
| 11 | rs2722769 | C | G | 0/7/156 | 0/4/166 | 0.38 |
| 11 | rs1387153 | A | G | 7/55/101 | 11/70/89 | 0.20 |
| 11 | rs7107217 | C | A | 39/73/46 | 33/90/45 | 0.37 |
| 12 | rs1153188 | A | T | 8/41/112 | 5/58/100 | 0.12 |
| 12 | rs1531343 | G | C | 36/67/57 | 31/67/68 | 0.55 |
| 12 | rs7961581 | G | A | 10/39/113 | 8/44/117 | 0.82 |
| 12 | rs7957197 | A | T | 14/67/82 | 10/73/87 | 0.64 |
| 13 | rs1359790 | A | G | 1/15/145 | 0/15/155 | 0.77 |
| 14 | rs730570 | A | G | 5/58/98 | 6/53/107 | 0.75 |
| 15 | rs7172432 | A | G | 1/45/115 | 4/46/118 | 0.51 |
| 15 | rs1436955 | G | A | 35/80/47 | 22/87/60 | 0.09 |
| 15 | rs7178572 | A | G | 27/82/54 | 22/86/62 | 0.60 |
| 15 | rs11634397 | G | A | 34/71/58 | 24/89/57 | 0.18 |
| 15 | rs2028299 | C | A | 15/74/74 | 14/66/90 | 0.38 |
| 15 | rs8042680 | C | A | 0/5/158 | 0/3/167 | 0.52 |
| 16 | rs8050136 | A | C | 46/58/59 | 41/85/44 | 0.02 |
| 16 | rs17797882 | A | G | 0/19/136 | 3/28/134 | 0.12 |
| 16 | rs16955379 | A | G | 3/55/105 | 6/48/116 | 0.37 |
| 17 | rs391300 | G | A | 19/76/68 | 12/77/81 | 0.28 |
| 17 | rs4430796 | A | G | 6/60/97 | 6/51/113 | 0.39 |

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 18 | rs10460009 | A | G | 1/20/142 | 1/17/152 | 0.81 |
| 19 | rs472265 | G | A | 21/85/56 | 20/91/59 | 0.96 |
| 20 | rs4812829 | A | G | 0/15/146 | 1/17/146 | 0.84 |
| 21 | rs2833610 | A | G | 17/82/64 | 21/72/76 | 0.37 |
| 23 | rs5945326 | G | A | 2/27/99 | 1/20/65 | 0.83 |

SNP=single nucleotide polymorphisms; A = adenine; G = guanine; C = cytosine; T = thymine; EMP1 = point-wise empirical value

# APPENDIX C

## Dominant genetic model univariate and multivariate results for the selected SNPs

This appendix contains the univariate and multivate test results for the dominant genetic model that are illustrated in Table C1.

**Table C1 Univariate test results for the dominant genetic model**

| CHROMOSOME | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 1 | rs7542900 | A | G | 110/53 | 111/57 | 0.82 |
| 1 | rs10923931 | A | C | 60/103 | 76/94 | 0.16 |
| 2 | rs7578597 | G | A | 75/88 | 67/102 | 0.27 |
| 2 | rs243021 | A | G | 121/42 | 121/49 | 0.55 |
| 2 | rs7560163 | C | G | 52/109 | 59/109 | 0.63 |
| 2 | rs3923113 | A | C | 44/119 | 48/122 | 0.82 |
| 2 | rs7578326 | G | A | 112/50 | 114/55 | 0.78 |
| 2 | rs2943641 | A | G | 85/64 | 86/65 | 0.98 |
| 3 | rs17036101 | A | G | 37/126 | 47/123 | 0.32 |
| 3 | rs358806 | A | C | 7/119 | 13/104 | 0.11 |
| 3 | rs831571 | A | G | 29/134 | 44/126 | 0.08 |
| 3 | rs4607103 | A | G | 91/72 | 100/70 | 0.66 |
| 3 | rs3773506 | G | C | 40/123 | 37/133 | 0.58 |
| 3 | rs7630877 | A | G | 93/48 | 92/48 | 0.93 |
| 3 | rs4402960 | C | A | 115/48 | 115/55 | 0.62 |
| 3 | rs16861329 | A | G | 6/157 | 8/162 | 0.79 |
| 4 | rs6815464 | G | C | 38/124 | 39/129 | 0.94 |
| 4 | rs4689388 | G | A | 87/76 | 90/79 | 1.00 |
| 5 | rs4457053 | G | A | 75/87 | 75/94 | 0.72 |
| 5 | rs12518099 | G | A | 79/83 | 90/76 | 0.38 |
| 6 | rs7754840 | G | C | 106/28 | 100/40 | 0.16 |
| 6 | rs9470794 | G | A | 71/92 | 78/92 | 0.69 |
| 6 | rs9472138 | A | G | 79/84 | 99/69 | 0.06 |
| 6 | rs1048886 | G | A | 86/77 | 86/84 | 0.75 |
| 7 | rs864745 | G | A | 49/114 | 46/123 | 0.63 |
| 7 | rs7636 | A | G | 83/80 | 85/84 | 0.91 |

| CHROMOSOME | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 7 | rs6467136 | A | G | 71/83 | 62/91 | 0.37 |
| 7 | rs972283 | A | G | 39/124 | 29/141 | 0.15 |
| 8 | rs896854 | G | A | 59/94 | 63/90 | 0.65 |
| 8 | rs13266634 | A | G | 28/126 | 24/138 | 0.42 |
| 9 | rs7041847 | G | A | 20/143 | 23/147 | 0.75 |
| 9 | rs17584499 | A | G | 23/140 | 24/138 | 0.88 |
| 9 | rs564398 | G | A | 5/158 | 2/167 | 0.29 |
| 9 | rs7018475 | C | A | 38/123 | 35/130 | 0.62 |
| 9 | rs13292136 | A | G | 14/149 | 14/156 | 1.00 |
| 10 | rs12779790 | G | A | 31/131 | 41/128 | 0.29 |
| 10 | rs6583826 | G | A | 114/42 | 127/37 | 0.36 |
| 10 | rs1111875 | A | G | 61/102 | 66/104 | 0.81 |
| 10 | rs7903146 | A | G | 76/85 | 89/79 | 0.31 |
| 10 | rs10741243 | G | C | 84/77 | 89/79 | 0.91 |
| 11 | rs231362 | A | G | 52/109 | 47/123 | 0.34 |
| 11 | rs2722769 | C | G | 7/156 | 4/166 | 0.34 |
| 11 | rs1387153 | A | G | 62/101 | 81/89 | 0.09 |
| 11 | rs7107217 | C | A | 112/46 | 123/45 | 0.62 |
| 12 | rs1153188 | A | T | 49/112 | 63/100 | 0.13 |
| 12 | rs1531343 | G | C | 103/57 | 98/68 | 0.31 |
| 12 | rs7961581 | G | A | 49/113 | 52/117 | 0.98 |
| 12 | rs7957197 | A | T | 81/82 | 83/87 | 0.91 |
| 13 | rs1359790 | A | G | 16/145 | 15/155 | 0.74 |
| 14 | rs730570 | A | G | 63/98 | 59/107 | 0.51 |
| 15 | rs7172432 | A | G | 46/115 | 50/118 | 0.83 |
| 15 | rs1436955 | G | A | 115/47 | 109/60 | 0.23 |
| 15 | rs7178572 | A | G | 109/54 | 108/62 | 0.57 |
| 15 | rs11634397 | G | A | 105/58 | 113/57 | 0.72 |
| 15 | rs2028299 | C | A | 89/74 | 80/90 | 0.19 |
| 15 | rs8042680 | C | A | 5/158 | 3/167 | 0.49 |
| 16 | rs8050136 | A | C | 104/59 | 126/44 | 0.04 |
| 16 | rs17797882 | A | G | 19/136 | 31/134 | 0.10 |
| 16 | rs16955379 | A | G | 58/105 | 54/116 | 0.49 |
| 17 | rs391300 | G | A | 95/68 | 89/81 | 0.34 |
| 17 | rs4430796 | A | G | 66/97 | 57/113 | 0.20 |
| 18 | rs10460009 | A | G | 21/142 | 18/152 | 0.60 |
| 19 | rs472265 | G | A | 106/56 | 111/59 | 0.95 |

| CHROMOSOME | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 20 | rs4812829 | A | G | 15/146 | 18/146 | 0.60 |
| 21 | rs2833610 | A | G | 99/64 | 93/76 | 0.29 |
| 23 | rs5945326 | G | A | 29/99 | 21/65 | 0.73 |

SNP=single nucleotide polymorphisms; A = adenine; G = guanine; C = cytosine; T = thymine; EMP1 = point-wise empirical value

# APPENDIX D
## Recessive genetic model univariate for the selected SNPs

This appendix contains the univariate test results for the recessive genetic model that are illustrated in Table D1.

**Table D1 Univariate test results of the recessive genetic model**

| CHROMOSOME | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 1 | rs7542900 | A | G | 24/139 | 42/126 | 0.02 |
| 1 | rs10923931 | A | C | 10/153 | 11/159 | 1.00 |
| 2 | rs7578597 | G | A | 8/155 | 7/162 | 0.80 |
| 2 | rs243021 | A | G | 27/136 | 34/136 | 0.48 |
| 2 | rs7560163 | C | G | 11/150 | 7/161 | 0.32 |
| 2 | rs3923113 | A | C | 5/158 | 3/167 | 0.49 |
| 2 | rs7578326 | G | A | 28/134 | 34/135 | 0.58 |
| 2 | rs2943641 | A | G | 19/130 | 14/137 | 0.36 |
| 3 | rs17036101 | A | G | 1/162 | 4/166 | 0.37 |
| 3 | rs358806 | A | C | 0/126 | 0/117 | 1.00 |
| 3 | rs831571 | A | G | 1/162 | 6/164 | 0.12 |
| 3 | rs4607103 | A | G | 19/144 | 26/144 | 0.33 |
| 3 | rs3773506 | G | C | 3/160 | 3/167 | 1.00 |
| 3 | rs7630877 | A | G | 20/121 | 18/122 | 0.78 |
| 3 | rs4402960 | C | A | 31/132 | 30/140 | 0.78 |
| 3 | rs16861329 | A | G | 0/163 | 0/170 | 1.00 |
| 4 | rs6815464 | G | C | 3/159 | 3/165 | 0.80 |
| 4 | rs4689388 | G | A | 11/152 | 9/160 | 0.64 |
| 5 | rs4457053 | G | A | 11/151 | 14/155 | 0.63 |
| 5 | rs12518099 | G | A | 7/155 | 12/154 | 0.27 |
| 6 | rs7754840 | G | C | 19/115 | 12/128 | 0.16 |
| 6 | rs9470794 | G | A | 13/150 | 13/157 | 1.00 |
| 6 | rs9472138 | A | G | 11/152 | 17/151 | 0.26 |
| 6 | rs1048886 | G | A | 11/152 | 14/156 | 0.69 |
| 7 | rs864745 | G | A | 4/159 | 4/165 | 1.00 |
| 7 | rs7636 | A | G | 20/143 | 21/148 | 1.00 |
| 7 | rs6467136 | A | G | 7/147 | 7/146 | 0.85 |
| 7 | rs972283 | A | G | 2/161 | 2/168 | 1.00 |

| CHROMOSOME | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 8 | rs896854 | G | A | 4/149 | 6/147 | 0.53 |
| 8 | rs13266634 | A | G | 4/150 | 9/153 | 0.23 |
| 9 | rs7041847 | G | A | 1/162 | 0/170 | 0.48 |
| 9 | rs17584499 | A | G | 1/162 | 0/162 | 0.65 |
| 9 | rs564398 | G | A | 0/163 | 0/169 | 1.00 |
| 9 | rs7018475 | C | A | 1/160 | 2/163 | 0.81 |
| 9 | rs13292136 | A | G | 2/161 | 0/170 | 0.23 |
| 10 | rs12779790 | G | A | 2/160 | 2/167 | 1.00 |
| 10 | rs6583826 | G | A | 28/128 | 38/126 | 0.27 |
| 10 | rs1111875 | A | G | 11/152 | 13/157 | 0.81 |
| 10 | rs7903146 | A | G | 15/146 | 12/156 | 0.51 |
| 10 | rs10741243 | G | C | 11/150 | 15/153 | 0.50 |
| 11 | rs231362 | A | G | 5/156 | 5/165 | 1.00 |
| 11 | rs2722769 | C | G | 0/163 | 0/170 | 1.00 |
| 11 | rs1387153 | A | G | 7/156 | 11/159 | 0.49 |
| 11 | rs7107217 | C | A | 39/119 | 33/135 | 0.26 |
| 12 | rs1153188 | A | T | 8/153 | 5/158 | 0.40 |
| 12 | rs1531343 | G | C | 36/124 | 31/135 | 0.40 |
| 12 | rs7961581 | G | A | 10/152 | 8/161 | 0.60 |
| 12 | rs7957197 | A | T | 14/149 | 10/160 | 0.39 |
| 13 | rs1359790 | A | G | 1/160 | 0/170 | 0.11 |
| 14 | rs730570 | A | G | 5/156 | 6/160 | 0.84 |
| 15 | rs7172432 | A | G | 1/160 | 4/164 | 0.31 |
| 15 | rs1436955 | G | A | 35/127 | 22/147 | 0.05 |
| 15 | rs7178572 | A | G | 27/136 | 22/148 | 0.39 |
| 15 | rs11634397 | G | A | 34/129 | 24/146 | 0.11 |
| 15 | rs2028299 | C | A | 15/148 | 14/156 | 0.84 |
| 15 | rs8042680 | C | A | 0/163 | 0/170 | 1.00 |
| 16 | rs8050136 | A | C | 46/117 | 41/129 | 0.44 |
| 16 | rs17797882 | A | G | 0/155 | 3/162 | 0.23 |
| 16 | rs16955379 | A | G | 3/160 | 6/164 | 0.50 |
| 17 | rs391300 | G | A | 19/144 | 12/158 | 0.19 |
| 17 | rs4430796 | A | G | 6/157 | 6/164 | 1.00 |
| 18 | rs10460009 | A | G | 1/162 | 1/169 | 1.00 |
| 19 | rs472265 | G | A | 21/141 | 20/150 | 0.74 |
| 20 | rs4812829 | A | G | 0/161 | 1/163 | 0.83 |
| 21 | rs2833610 | A | G | 17/146 | 21/148 | 0.54 |

| CHROMOSOME | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 23 | rs5945326 | G | A | 2/126 | 1/85 | 0.59 |

SNP=single nucleotide polymorphisms; A = adenine; G = guanine; C = cytosine; T = thymine; EMP1 = point-wise empirical value

# APPENDIX E

## CATT genetic model univariate results for the selected SNPs

This appendix contains the univariate test results for the additive genetic model that are illustrated in Table E1.

**Table E1 Univariate test results of the CATT genetic model**

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 1 | rs7542900 | A | G | 134/192 | 153/183 | 0.27 |
| 1 | rs10923931 | A | C | 70/256 | 87/253 | 0.24 |
| 2 | rs7578597 | G | A | 83/243 | 74/264 | 0.28 |
| 2 | rs243021 | A | G | 148/178 | 155/185 | 1.00 |
| 2 | rs7560163 | C | G | 63/259 | 66/270 | 0.97 |
| 2 | rs3923113 | A | C | 49/277 | 51/289 | 0.99 |
| 2 | rs7578326 | G | A | 140/184 | 148/190 | 0.91 |
| 2 | rs2943641 | A | G | 104/194 | 100/202 | 0.63 |
| 3 | rs17036101 | A | G | 38/288 | 51/289 | 0.21 |
| 3 | rs358806 | A | C | 7/245 | 13/221 | 0.15 |
| 3 | rs831571 | A | G | 30/296 | 50/290 | 0.03 |
| 3 | rs4607103 | A | G | 110/216 | 126/214 | 0.38 |
| 3 | rs3773506 | G | C | 43/283 | 40/300 | 0.62 |
| 3 | rs7630877 | A | G | 113/169 | 110/170 | 0.84 |
| 3 | rs4402960 | C | A | 146/180 | 145/195 | 0.54 |
| 3 | rs16861329 | A | G | 6/320 | 8/332 | 0.78 |
| 4 | rs6815464 | G | C | 41/283 | 42/294 | 0.93 |
| 4 | rs4689388 | G | A | 98/228 | 99/239 | 0.79 |
| 5 | rs4457053 | G | A | 86/238 | 89/249 | 0.93 |
| 5 | rs12518099 | G | A | 86/238 | 102/230 | 0.23 |
| 6 | rs7754840 | G | C | 125/143 | 112/168 | 0.06 |
| 6 | rs9470794 | G | A | 84/242 | 91/249 | 0.77 |
| 6 | rs9472138 | A | G | 90/236 | 116/220 | 0.05 |
| 6 | rs1048886 | G | A | 97/229 | 100/240 | 0.87 |
| 7 | rs864745 | G | A | 53/273 | 50/288 | 0.61 |

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 7 | rs7636 | A | G | 103/223 | 106/232 | 0.97 |
| 7 | rs6467136 | A | G | 78/230 | 69/237 | 0.44 |
| 7 | rs972283 | A | G | 41/285 | 31/309 | 0.17 |
| 8 | rs896854 | G | A | 63/243 | 69/237 | 0.54 |
| 8 | rs13266634 | A | G | 32/276 | 33/291 | 0.95 |
| 9 | rs7041847 | G | A | 21/305 | 23/317 | 0.77 |
| 9 | rs17584499 | A | G | 24/302 | 24/300 | 0.99 |
| 9 | rs564398 | G | A | 5/321 | 2/336 | 0.28 |
| 9 | rs7018475 | C | A | 39/283 | 37/293 | 0.73 |
| 9 | rs13292136 | A | G | 16/310 | 14/326 | 0.63 |
| 10 | rs12779790 | G | A | 33/291 | 43/295 | 0.32 |
| 10 | rs6583826 | G | A | 142/170 | 165/163 | 0.20 |
| 10 | rs1111875 | A | G | 72/254 | 79/261 | 0.76 |
| 10 | rs7903146 | A | G | 91/231 | 101/235 | 0.62 |
| 10 | rs10741243 | G | C | 95/227 | 104/232 | 0.70 |
| 11 | rs231362 | A | G | 57/265 | 52/288 | 0.40 |
| 11 | rs2722769 | C | G | 7/319 | 4/336 | 0.36 |
| 11 | rs1387153 | A | G | 69/257 | 92/248 | 0.07 |
| 11 | rs7107217 | C | A | 151/165 | 156/180 | 0.74 |
| 12 | rs1153188 | A | T | 57/265 | 68/258 | 0.26 |
| 12 | rs1531343 | G | C | 139/181 | 129/203 | 0.27 |
| 12 | rs7961581 | G | A | 59/265 | 60/278 | 0.90 |
| 12 | rs7957197 | A | T | 95/231 | 93/247 | 0.64 |
| 13 | rs1359790 | A | G | 17/305 | 15/325 | 0.63 |
| 14 | rs730570 | A | G | 68/254 | 65/267 | 0.61 |
| 15 | rs7172432 | A | G | 47/275 | 54/282 | 0.61 |
| 15 | rs1436955 | G | A | 150/174 | 131/207 | 0.04 |
| 15 | rs7178572 | A | G | 136/190 | 130/210 | 0.41 |
| 15 | rs11634397 | G | A | 139/187 | 137/203 | 0.59 |
| 15 | rs2028299 | C | A | 104/222 | 94/246 | 0.20 |
| 15 | rs8042680 | C | A | 5/321 | 3/337 | 0.51 |
| 16 | rs8050136 | A | C | 150/176 | 167/173 | 0.50 |
| 16 | rs17797882 | A | G | 19/291 | 34/296 | 0.06 |
| 16 | rs16955379 | A | G | 61/265 | 60/280 | 0.73 |
| 17 | rs391300 | G | A | 114/212 | 101/239 | 0.15 |
| 17 | rs4430796 | A | G | 72/254 | 63/277 | 0.25 |

| CHROMOSOME NUMBER | SNP | MINOR ALLELE | MAJOR ALLELE | CASES | CONTROLS | EMP1 |
|---|---|---|---|---|---|---|
| 18 | rs10460009 | A | G | 22/304 | 19/321 | 0.61 |
| 19 | rs472265 | G | A | 127/197 | 131/209 | 0.84 |
| 20 | rs4812829 | A | G | 15/307 | 19/309 | 0.53 |
| 21 | rs2833610 | A | G | 116/210 | 114/224 | 0.61 |
| 23 | rs5945326 | G | A | 31/225 | 22/150 | 0.84 |

SNP=single nucleotide polymorphisms; A = adenine; G = guanine; C = cytosine; T = thymine; EMP1 = point-wise empirical value