

Genetic and dietary determinants of Type 2 diabetes in a black South African population

T Chikowore
23917245

Thesis submitted in *partial* fulfillment of the requirements for the degree *Doctor Philosophiae* in Nutrition at the Potchefstroom Campus of the North-West University

Supervisor: Dr K R Conradie
Co-supervisor: Dr T Van Zyl
Co-supervisor: Prof E Feskens

November 2016

Acknowledgements

To the Lord who ignited the desire, gave the ability, ensured my health and granted the resources be glory and honour forever AMEN. I would to like to thank HIM for uniting me with these people that made this work come to fruition.

I would immensely want to thank **Dr KR Conradie** for believing in me and granting the resources and emotional support such that I could grow academically and complete this work. Thank you exceedingly for your commitment to ensure that I could excel and the recommendations you gave that opened many doors of opportunity and awards. Indeed it takes great people like you to inspire excellence in young and timid people like me.

I would like to thank **Dr T van Zyl** for being there and ever willing to guide me along this path. I admired your genuine remarks and learnt a lot from you. Thank you for being a shoulder where I could echo my burning emotions and above all thank you being there during the trying times.

I am grateful to **Prof EJM Feskens** for your generosity that helped us perform the novel work of sequencing the whole genomes of the Setswana people for the first time. Thank you so much for your valued experience you brought into the study, that made us obtain significant outputs.

I would like to thank **Dr P Pisa** for your belief in me and much more for being there to ignite this dream and making sure, I come to Potchefstroom to obtain it. I am grateful. Much thanks to our collaborators, **Prof A Morris, Prof M Ramsay, Prof E Viljoen, Prof S Hazelhurst and Shaun Aron** for the ideas and support which helped us implement and write outputs of the study.

Special thanks goes to my wife **Priscilla Masawi** and family for the emotional support and sacrifices you made to help me complete this work. I also acknowledge the PURE research team, participants and funders. I am grateful to the NRF for the financial support. Above all I thank the Lord for once again leading me thus far and to HIM be glory, forever, AMEN.

Abstract

Background: The increasing burden of T2D is a global cause of concern. However in Africa where the countries are ill equipped to manage this pandemic, the prevalence of T2D will increase the most by 2030 compared to other parts of the world. Urgent efforts are thus required to lessen the T2D pandemic. It is postulated that adoption of Westernised lifestyles among people who have a genetic predisposition to T2D is fuelling this epidemic. However, it is not clear which nutrient factors and genetic factors interact to increase T2D risk in the black South African population. We set out to determine 1) the predictive utility of the European and Asian derived variants associated with T2D in a black South African population; 2) the rare and low frequency variants associated with T2D using novel whole genome sequences of people of Setswana descent; 3) the nutrient patterns associated with glycated hemoglobin and fasting glucose in a black South African population.

Methods: Four types of genetic risk scores (GRS) were computed using the 66 SNPs that had been associated with T2D in the Europeans and Asians. This comprised of the GRSt which consisted of all the 66 variants; GRSn that comprised of the variants associated with T2D in the black South African population of Sestwana descent; GRSb comprised of variants that had been characterised to be involved in the functioning of β cells and GRStrans which comprised of variants that were associated with T2D amongst varied ethnicities. Receiver operating curves (ROC) were used to evaluate the predictive utility of the GRS that was found to be associated with T2D. Extreme phenotype sampling was used to select 16 T2D cases and 14 controls from the tails of the glycated haemoglobin distribution after the adjustment of age, sex, BMI and fasting glucose for whole genome sequencing. The EPACTS software was used to conduct the single variant and gene-based association test using EMMAX to adjust for population structure using whole genome sequence data. Principal component analysis was applied onto 25 nutrients to determine the nutrient patterns in the study population. The associations of the extracted nutrient patterns with fasting glucose and glycated hemoglobin were evaluated.

Results: The GRSn was the only polygenic risk score which was significantly associated with T2D in the study population. It was noted to modestly improve the predictive utility of

the known risk factors by 2%. However, it was noted to be more predictive of risk among people who were lean and less than 50 years of age. The variant rs7499753 was observed to be associated with T2D beyond the genome wide significance level. However, replication studies are required to confirm this association. Starch, dietary fibre and B vitamins driven nutrient pattern was significantly associated with fasting glucose ($\beta = -0.224$ (-0.446; -0.002); $p = 0.048$), and glycated hemoglobin levels ($\beta = -0.162$ (-0.291; -0.033); $p = 0.014$) in rural women. Thiamine, zinc and plant protein driven nutrient pattern was associated with significant reductions in fasting glucose and glycated hemoglobin of ($\beta = -0.435$ (-0.802; -0.067) $p = 0.021$) and ($\beta = -0.301$ (-0.554; -0.049); $p = 0.020$) in rural men.

Conclusion: The GRS comprising of variants associated with T2D risk in the Setswana population improved modestly (by 2%) the prediction of this disease in addition to conventional risk factors. More population specific variants are required to develop a GRS that is more predictive of the T2D risk in the Setswana population. A new association of rs7499753 was documented that still needs to be confirmed through replication studies. More, well powered studies are required to whole genome sequence association studies through collaborative initiatives to elucidate more of the population specific variants associated with T2D among the Setswana population group. The consumption of plant based nutrients was noted to be associated with decreases in fasting and glycated hemoglobin. Clinical trials consisting of iso-caloric diets are required to further explore the nutrient patterns associated with T2D risk.

TABLE OF CONTENTS

Acknowledgements	i
Abstract	ii
List of Tables	viii
List of Figures	ix
List of Abbreviations	x
CHAPTER 1	1
Introduction	1
1.1 Aims and Objectives	3
1.1.1 Aim:	3
1.1.2 Objectives.....	3
1.2 Research Team and contributions.....	4
1.3 Outline of the thesis	5
1.4 References.	6
CHAPTER 2	8
Literature Review.....	8
2.2 Diagnosis of T2D	9
2.3 Pathophysiology of T2D	10
2.3.1 Insulin resistance vs β -cell dysfunction	12
2.3.2 Intrauterine environment, infancy and T2D risk	13
2.3.3 Environmental chemicals.....	14

2.3.4	Microbiome	14
2.4	Genetics of T2D	14
2.4.1	Pre-GWAS era	15
2.4.2	GWAS of T2D in European populations	16
2.4.3	GWAS of T2D among Asians	18
2.4.4	Trans-ethnic GWAS of T2D	18
2.4.5	GWAS and missing heritability	19
2.4.6	Extreme phenotype sampling and WGS.....	25
2.4.7	Polygenic risk scores and T2D	26
2.4.8	GWAS of T2D in people of Africa Ancestry	27
2.4.9	Genetic diversity among Africans	28
2.4.10	Genetics of T2D in South Africa	29
2.5	Nutrition and T2D	30
2.5.1	Nutrition transition.....	30
2.5.2	Macronutrients and T2D	31
2.5.3	Vitamins and minerals and T2D	31
2.5.4	Individual foods and T2D	32
2.5.5	Dietary patterns and T2D	32
2.5.6	Dietary patterns and T2D risk in black South Africans.....	34
2.5.7	Nutrient patterns vs Dietary Patterns.....	34
2.6	Summary	35
2.7	References	35

CHAPTER 3	54
The predictive utility of a genetic risk score of common variants associated with T2D in a black South African population	54
CHAPTER 4	80
Exploration of the genetic variants associated with type 2 diabetes using novel whole genome sequences in individuals of Setswana descent in South Africa	80
CHAPTER 5	104
Nutrient patterns associated with fasting and glycated hemoglobin in a black South African population.....	104
CHAPTER 6	136
Concluding Remarks.....	136
6.2 Nutrient patterns and t2d susceptibility	137
6.3 evaluation of the Genetic determinants of t2d using novel whole genome sequecnes of sestwana people	139
6.4 Implications of future research and recommendations	140
6.4.1 Polygenetic risk scores and T2D	141
6.4.2 Nutrient patterns and T2D	141
6.4.3 Rare variants and T2D risk.....	141
6.5 References	142
APPENDIX A.....	145
Ethical Approval and Consent forms.....	145

APPENDIX B.....	153
Published Manuscript	153
APPENDIX C.....	162
Acceptance letter for publication: Nutrient patterns associated with fasting glucose and glycated hemoglobin in a black South African population.	162

List of Tables

Table 1.1	Contributions of the research team members	4
Table 2.1	Diagnostic criteria for diabetes	10
Table 2.2	List of common variants that have been discovered through GWAS	19
Table 2.3	Dietary patterns and their associations with T2D risk.....	33

List of Figures

Figure 2.1	Schematic diagram of the pathophysiology of T2D	11
Figure 2.2	Aetiological framework for T2D	13

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
AGEs	Advanced glycation end products
AGVP	African Human Genetic Variation Programme
AP3S2	adaptor-related protein complex 3, sigma 2 subunit
ARAP1	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1
ASO	Allele specific oligonucleotides
AUC	Area under the curve
BARX2	BARX homeobox 2
BCL11A	B-cell CLL/lymphoma 11A (zinc finger protein)
BPA	Bisphenol
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
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
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
EMMAX	Efficient mixed model expediated
F3	coagulation factor III (thromboplastin, tissue factor)
FPG	Fasting plasma glucose

FTO	Fat mass and obesity-associated
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
GLUT4	Glucose transporter type 4
GRB14	Growth factor receptor-bound protein 14
GRS	Genetic risk score
GWAS	Genome-wide association studies
H3 Africa	Human, Heredity and Health in Africa
HbA1C	Glycated haemoglobin
HHEX	Haematopoietically expressed homeobox
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
PC	Principal component
PCA	Principal component analysis
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
QC	Quality control
QFFQ	Quantitative food frequency questionnaire
RBM43	RNA binding motif protein 43
RBMS1	RNA binding motif, single stranded interacting protein 1
RND3	Rho family GTPase 3

ROC	Receiver operating curves
SEMDSA	Society for Endocrinology, Metabolism and Diabetes of South Africa
SKAT-O	Sequence association test -optimised
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-sialyltransferase 1
SYN2	Synapsin II
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
TP53NP1	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 (<i>S. pombe</i>)
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

CHAPTER 1

Introduction

Type 2 diabetes (T2D) is a complex disease characterised hyperglycaemia and altered lipid metabolism which results in the damage of body organs leading to conditions such as blindness, heart diseases and kidney failure. The increasing prevalence of T2D is a major public health concern. It has been estimated by the International Diabetes Federation that the number of people with T2D will increase from 366 million in 2011 to 552 million in 2030 (Sanghera & Blackett, 2012). However, Africa which currently is ill-equipped to manage the T2D burden, is postulated to experience the largest worldwide increase in T2D prevalence of 109% by 2030. The black South African population has not been spared by the increasing prevalence and burden of T2D. The highest age-standardised prevalence of T2D in Sub Saharan Africa (13.1%) was reported among the black South Africans of Cape Town (Peer *et al.*, 2012). Coupled with the increase in T2D prevalence is also the increase in the economic burden of this disease. Projected increases in the health care costs of diabetics in Sub Saharan Africa by 50% from 2010 to 2030 pose a great economic threat to countries in these region if drastic measures to curb this pandemic are not implemented. In South Africa these costs are estimated to be about 2 billion USD in 2030. This growing pandemic is attributed to the adoption of Westernised diets and lifestyles among people who are genetically susceptible to T2D (Franks *et al.*, 2013). However, there is no clear evidence of the precise genetic and dietary determinants associated with increased T2D risk in the black South African setting.

The genetic aetiology of T2D has been extensively evaluated through genome-wide association studies (GWAS) among Asian and European ethnicities but no GWAS for T2D has been conducted in Africa (Florez, 2013). The results of GWAS have been disappointing for T2D, as the associated variants have been noted to explain only about 10% of the heritability of this disease (Sanghera & Blackett, 2012). However, in a number

of studies, it has been determined that a combination of common variants computed into a genetic risk score (GRS) has a higher predictive ability to denote T2D risk compared with individual SNPs (Weedon *et al.*, 2006; Lango *et al.*, 2008). Notably, in a study by Vassy *et al.* 2014, a similar GRS was observed to be associated with T2D risk among Caucasians and African Americans, thereby suggesting that this approach is not limited by genetic differences in these population groups (Vassy *et al.*, 2014). It is therefore important to evaluate whether the GRS composed of these European-derived SNPs is also associated with T2D risk among other African population groups. The genetic risk score has been depicted as more predictive in participants who are non-obese and less than 50 years old (de Miguel-Yanes *et al.*, 2011). Thus, the GRS is relevant for identifying high-risk individuals while they are still young and have not yet developed risk factors such as obesity as most people when they diagnosed would have developed irreversible complications. However, the GRS approach has not been used to assess the disease risk predicted by the reported common variants associated with T2D risk in the black South African population. Therefore a need exists to evaluate the association of the joint effects of common variants in the form of GRS with T2D risk and in the black South African population. The inclusion or rare variants that are postulated to have large effects is pivotal to enhance the predictivity of the GRS associated with T2D.

The advances in next-generation whole-genome sequencing technologies are thought to enable the identification of these rare variants associated with T2D (Zhang *et al.*, 2011). Recent advances in technology have reduced the costs of whole-genome sequencing and it has now become accessible to scientists in low-income countries (Sargolzaei *et al.*, 2014). Whole-genome sequences are pivotal in newly studied ethnic populations as they can allow detection of population-specific common and rare variants, since European-derived variants cannot be used to extrapolate SNPs associated with disease risk among African ethnicities (Sanghera & Blackett, 2012; Mhandire *et al.*, 2014). Therefore, it is crucial to use whole-genome sequence data to identify population-specific variants associated with T2D in the black South African population. However, although T2D has a significant genetic component as part of its aetiology, the recent changes in the eating patterns are also fuelling the prevalence of this disease.

A number of studies have reported that the nutrition transition from traditional to Western eating patterns is exacerbating the prevalence of chronic diseases, which include T2D in the black South African population (MacIntyre *et al.*, 2002; Vorster *et al.*, 2011). However, limited evidence exists of the precise dietary components that are associated with increased T2D risk as the nutrition transition is generalised for increasing disease risk of a number of lifestyle-related NCDs, such as hypertension. Conversely, the urban migration has been linked to improvements in micronutrient intake owing to improved access to fruit and vegetables among urban women (Vorster *et al.*, 2005; Hattingh *et al.*, 2008). This makes the association of dietary factors with T2D more complex, as both fruits and vegetables are protective against T2D, while intake of animal protein, sugar and fat increases the risk of this disease. Added to this, people eat meals with a variety of nutrients which have interactive and synergistic effects on health (Hu, 2002). Therefore, it is difficult to determine the separate effect of a food or nutrient on disease development as it is highly interrelated with other nutrients (Hu, 2002). There is a need for nutrient pattern analysis methods which are able to evaluate the diet as a whole and clarify the effects of how the consumption of sugary foods and meat products, together with improved intakes of fruits and vegetables, are related to T2D risk amongst this population group.

1.1 AIMS AND OBJECTIVES

1.1.1 Aim:

- To determine the genetic and dietary factors associated with T2D susceptibility in a black South African population.

1.1.2 Objectives

- To determine the joint effect of common variants in the form of genetic risk score (GRS) associated with T2D susceptibility.

- To determine population specific rare and common variants from whole genome sequence data associated with T2D risk
- To determine the nutrient patterns associated with T2D susceptibility

1.2 RESEARCH TEAM AND CONTRIBUTIONS

Table 1.1 Contributions of the research team members

Team Member	Affiliation	Contribution
Mr T Chikowore	Centre of Excellence for Nutrition, North-West University	Implemented the extreme phenotype sampling and DNA sample preparation for the whole-genome sequencing and genotyping variants associated with T2D. Conceptualised and led out in the data analysis and writing up of all the manuscripts in this thesis.
Dr T Van Zyl (Co-promoter)	Centre of Excellence for Nutrition, North-West University	Assisted in securing ethical approval of the whole-genome sequencing, securing funds for the genotyping of T2D associated SNPs and writing of the manuscripts. Helped with securing the data for the dietary analysis.
Prof. EJM Feskens (Co-promoter)	Wageningen University, Division of Human Nutrition, P.O. Box 17, 6700 AA Wageningen, The Netherlands	Assisted in securing additional funding for whole-genome sequencing and in the writing of the manuscripts

Dr KR Conradie (Promoter)	Centre of Excellence for Nutrition, North-West University	Initiated, partly conceptualised, planned and secured funding for the whole- genome sequencing analysis and genotyping of T2D associated variants. Assisted also with the writing of the manuscripts.
------------------------------	---	--

1.3 OUTLINE OF THE THESIS

Chapter 1 indicates the rationale and objectives of the study and includes the study team. Chapter 2 includes the literature review which illustrates the current genetic factors associated with T2D in European and Asian ethnicities, highlighting the need for more extensive genetic studies among people of African ancestry. Limited evidence regarding the nutrient patterns associated with T2D is also illustrated in this chapter. Chapters 3 to 5 are formatted in line with the respective journal requirements to which the manuscripts were submitted.

Chapter 3 is a manuscript entitled “Predictive utility of common variants associated with T2D risk in a black South African population”. This manuscript was accepted for publication in the journal of *Diabetes Research and Clinical Practice* in September 2016. Impact factor 3.045 (See Appendix B)

Chapter 4 is a manuscript titled “Nutrient patterns associated with fasting glucose and glycated haemoglobin in a black South African population.” This manuscript was accepted for publication in the *Nutrients* journal. Impact factor 3.759 (See Appendix C)

Chapter 5 consists of a manuscript entitled “Exploration of the genetic determinants of type 2 diabetes among novel whole-genome sequences in individuals of Setswana descent”. This manuscript was submitted to *Plos Genetics*. Impact factor 6.661

In Chapter 6 the conclusions, limitations of the study and future perspectives are discussed.

1.4 REFERENCES.

de Miguel-Yanes, J.M., Shrader, P., Pencina, M.J., Fox, C.S., Manning, A.K., Grant, R.W., Dupuis, J., Florez, J.C., D'Agostino, R.B., Sr., Cupples, L.A., Meigs, J.B., Investigators, M. & Investigators, D. 2011. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes Care*, 34(1):121-125.

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., 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.

Hattingh, Z., Walsh, C.M., Bester, C.J. & Oguntibeju, O.O. 2008. Evaluation of energy and macronutrient intake of black women in Bloemfontein: A cross-sectional study. *African Journal of Biotechnology*, 7(22):4019-4024.

Hu, F.B. 2002. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*, 13(1):3-9.

Lango, H., Palmer, C.N.A., Morris, A.D., Zeggini, E., Hattersley, A.T., McCarthy, M.I., Frayling, T.M., Weedon, M.N. & Consortium, U.T.D.G. 2008. Assessing the Combined Impact of 18 Common Genetic Variants of Modest Effect Sizes on Type 2 Diabetes Risk. *Diabetes*, 57(11):3129-3135.

MacIntyre, U.E., Kruger, H.S., Venter, C.S. & Vorster, H.H. 2002. Dietary intakes of an African population in different stages of transition in the North West Province, South Africa: the THUSA study. *Nutrition Research*, 22(3):239-256.

Mhandire, K., Pharo, G., Kandawasvika, G.Q., Duri, K., Swart, M., Stray-Pedersen, B. & Dandara, C. 2014. How does mother-to-child transmission of HIV differ among African populations? Lessons from MBL2 genetic variation in Zimbabweans. *OMICS*, 18(7):454-460.

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.

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

Sargolzaei, M., Chesnais, J.P. & Schenkel, F.S. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15.

Vassy, J.L., Hivert, M.F., Porneala, B., Dauriz, M., Florez, J.C., Dupuis, J., Siscovick, D.S., Fornage, M., Rasmussen-Torvik, L.J., Bouchard, C. & Meigs, J.B. 2014. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes*, 63(6):2172-2182.

Vorster, H.H., Kruger, A. & Margetts, B.M. 2011. The Nutrition Transition in Africa: Can It Be Steered into a More Positive Direction? *Nutrients*, 3(4):429-441.

Vorster, H.H., Margetts, B.M., Venter, C.S. & Wissing, M.P. 2005. Integrated nutrition science: from theory to practice in South Africa. *Public Health Nutr*, 8(6A):760-765.

Weedon, M.N., McCarthy, M.I., Hitman, G., Walker, M., Groves, C.J., Zeggini, E., Rayner, N.W., Shields, B., Owen, K.R., Hattersley, A.T. & Frayling, T.M. 2006. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Medicine*, 3(10):1877-1882.

Zhang, J., Chiodini, R., Badr, A. & Zhang, G. 2011. The impact of next-generation sequencing on genomics. *J Genet Genomics*, 38(3):95-109.

CHAPTER 2

Literature Review

The prevalence of T2D is increasing rapidly and is now a major public health. However, the aetiology of this complex disease has yet to be clarified and this is hampering efforts to develop effective preventative and curative initiatives to curb its growing burden (Nolan *et al.*, 2011). It is postulated that changes in dietary patterns and lifestyles as a result of urbanisation among people with a genetic predisposition to T2D, are fuelling the increase in the prevalence of this disease burden (Franks *et al.*, 2013). This literature review explores current evidence regarding this phenomenon, indicating controversies, gaps in knowledge and the current context of the black South African population with regard to the genetic and dietary determinants of T2D.

2.1 DIABETES

T2D is a complex and multifactorial disease which is characterised by hyperglycaemia and altered lipid metabolism as a result of failure of β cells to secrete insulin in response to insulin resistance, inactivity, over-nutrition and obesity (Nolan *et al.*, 2011). It causes damage to body organs, leading to blindness, amputations and kidney failure. Other diseases, which include cardiovascular diseases, tuberculosis, alcoholic fatty liver disease and polycystic ovarian syndrome, are also associated with T2D (Alberti & Zimmet, 1998; Nolan *et al.*, 2011).

The symptoms of T2D consist of ketoacidosis, blurred vision, thirst and polyuria. However, it is postulated that T2D may begin and progress for 10 years before these signs and symptoms are visible and of concern to the patient (Saudek *et al.*, 2008). Therefore, it is estimated that 30-80% of diabetes cases are usually not diagnosed early and 25% of patients who are diagnosed would have already developed microvascular

complications (Harris, 1993; Saudek *et al.*, 2008). Therefore, strategies for enabling the early detection and diagnosis of T2D are required to curb this burden.

2.2 DIAGNOSIS OF T2D

Fasting plasma glucose, random plasma glucose, oral glucose tolerance test (OGTT) and glycated haemoglobin are biomarkers and methods used to diagnose T2D (Nolan *et al.*, 2011). The World Health Organisation (WHO) and the American Diabetes Association (ADA) reviewed the diagnostic criteria for T2D and suggested the cut-off values indicated in Table 2.1, which have been adopted globally (Nolan *et al.*, 2011). Fasting plasma glucose indicates the blood glucose levels after having fasted for at least 8 hours. According to the ADA 2013 guidelines, fasting plasma glucose levels of 5.6 to 6.9 mmol/l indicate impaired fasting glucose and levels of 7 mmol/l and higher indicate diabetes (American Diabetes, 2013). Random plasma glucose is used to assess the blood glucose levels without prior fasting. Through this method, diabetes is diagnosed when the levels are above 11.1 mmol/l and diabetes symptoms are present (Nolan *et al.*, 2011).

The OGTT involves giving the patient a predefined quantity of glucose (usually 75 grams) followed by measuring blood glucose concentrations after 30 minutes for a 2-hour period (Nolan *et al.*, 2011). The 2-hour post-prandial glucose level is used for diagnostic purposes and diabetes is considered if this measure is greater than 11.1 mmol/l (American Diabetes, 2013). Glycated haemoglobin is formed non-enzymatically through a glycation pathway as a result of the exposure of haemoglobin to glucose (WHO, 2011). It is indicative of the average plasma glucose levels over a longer period ranging from 2-3 months (WHO, 2011). The WHO and ADA both recommend glycated haemoglobin levels 6.5% higher for the diagnosis of diabetes (WHO, 2011; American Diabetes, 2013). However, countries such as South Africa in which iron deficiency is common have yet to adopt the use of glycated haemoglobin for diagnosis of diabetes, as they are still reviewing its relevance in view of the factors such as iron deficiency which are prevalent in this context and that might affect the accuracy of this approach (Amod A & Guideline, 2012).

Table 2.1 Diagnostic criteria for diabetes

Diagnostic method	Diabetes (WHO and ADA)	IFG and IGT (WHO)	Prediabetes (ADA)
Glycated haemoglobin (%)	≥ 6.5	NA	≥5.7 and < 6.5
Fasting Plasma glucose (mmol/l)	≥ 7.0	IFG ≥ 6.1 and < 7.0	≥ 5.6 and <7.0
OGTT (mmol/l)	≥ 11.1	IGT ≥ 7.8 and < 11.1	≥ 7.8 and < 11.1
Random glucose (mmol/l)	≥ 11.1 with classic symptoms	NA	NA

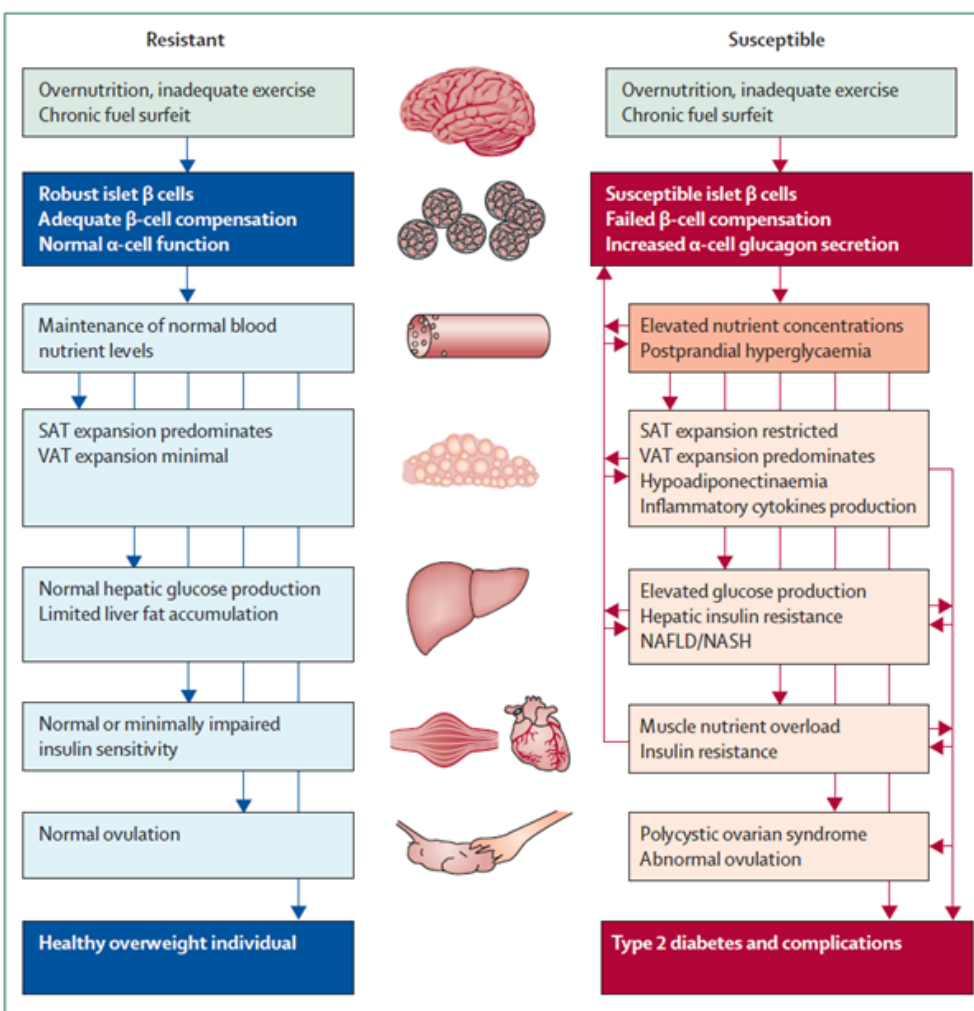
Adopted from (Nolan *et al.*, 2011). OGTT = oral glucose tolerance test; mmolL⁻¹ = millimoles per litre; IFG = impaired fasting glucose; IGT = impaired glucose tolerance; ADA = American Dietetic Association; NA = non-applicable; WHO = World Health Organisation; % = percentage

2.3 PATHOPHYSIOLOGY OF T2D

The development of the insulin radioimmunoassay enhanced the investigation of the pathophysiology of T2D (Yalow & Berson, 1960). It enabled the production of insulin in response to nutrient intake to be observed among diabetics (Yalow & Berson, 1960). In these earlier times it was noted that diabetic individuals were also not able to respond well to insulin and were regarded as “insulin insensitive” (Reaven, 1988). This insulin insensitivity was observed to lead to an increase in glucose production by the liver and a decrease in glucose uptake by the muscle and adipose tissues (Reaven, 1988). These abnormalities in the metabolism of glucose are now known to be a result of increased adiposity in the intra-abdominal cavity, which is regarded as central obesity (Cnop *et al.*, 2002).

Although obesity is a major risk factor for developing T2D, not everyone who is obese develops T2D (Nolan *et al.*, 2011). It is now known that diabetes-resistant individuals may be exposed to excess nutrients but are able to contain them by having active β cells which will release adequate insulin regardless of them being obese (Nolan *et al.*, 2011). Among these people, the increase in body size is due to the predominate increase in the protective subcutaneous adipose tissue (SAT) as opposed to the visceral adipose tissue (VAT) which is a risk of T2D, as indicated in Figure 2.1 (Nolan *et al.*, 2011). Thus the

nutrient levels are maintained in such a way that they will not cause damage to body tissue such as the liver, heart, ovaries and skeletal muscles, as illustrated in Figure 2.1. Conversely, obese people with a predisposition to T2D, when they are exposed to excess nutrients, are not able to control them, owing to β cells that are susceptible to failure when they are overworked. These people then experience an increase in VAT as opposed the protective SAT (Nolan *et al.*, 2011). Thus, inflammatory cytokines are raised in plasma leading to the damage of body tissue such as the liver, skeletal muscles, heart and ovaries. T2D then develops, further worsening tissue damage, as indicated in Figure 2.1.



Adopted from (Nolan *et al.*, 2011)

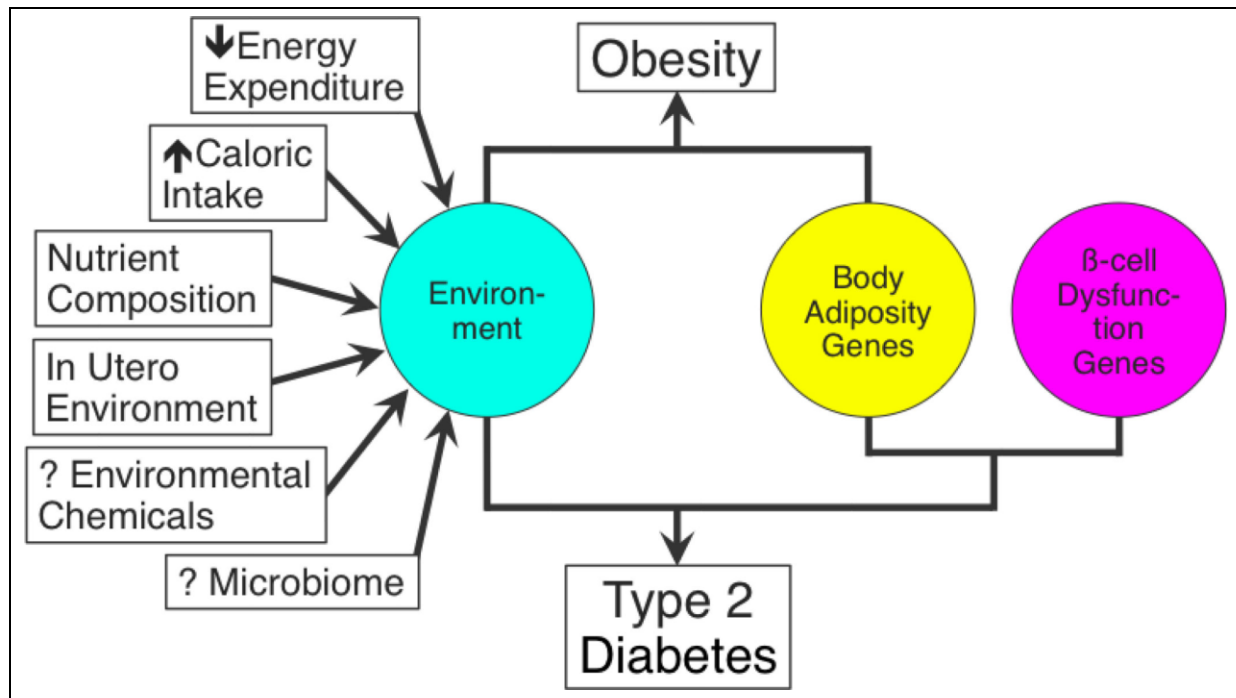
Figure 2.1 Schematic diagram of the pathophysiology of T2D

2.3.1 Insulin resistance vs β -cell dysfunction

A controversy regarding the initial defect in the pathogenesis of T2D between, insulin resistance and β -cell dysfunction, once existed (Reaven, 1988). Some scholars had the view that insulin resistance was the primary abnormality and that β -cell dysfunction came later on (Reaven, 1988). However, this view later changed with the realisation that a feedback mechanism existed between the β cells and the insulin-sensitive tissue (Kahn *et al.*, 1993). To the extent, that insulin resistance becomes much more pronounced due to the failure to release adequate insulin to prevent elevation of glucose as a result of β cell failure. Thus, insulin resistance and β -cell failure are primary determinants in the pathogenesis of T2D (Reaven, 1988).

The progressive decline of β -cell function has been shown to account for the change from impaired glucose tolerance to T2D (Weyer *et al.*, 1999). The existence of diminished β -cell function has been documented in first-degree relatives of diabetics, and in women with gestational diabetes or polycystic ovary syndrome (Cnop *et al.*, 2007). In addition, it has been indicated that β -cell function is heritable and critical in determining glucose intolerance and T2D in different racial and ethnic populations. Thus, β -cell dysfunction is now agreed to be just as important as insulin resistance in the pathogenesis of T2D.

However, despite the understanding of the role of β -cell dysfunction and insulin resistance in the aetiology of T2D, this disease is heterogeneous in nature, involving the interaction of many factors, illustrated in Figure 2.2, which are further explored in the coming sections.



Adopted from (Kahn *et al.*, 2014)

Figure 2.2 Aetiological framework for T2D

2.3.2 Intrauterine environment, infancy and T2D risk

The risk of T2D is postulated to begin during the foetal stages of life. According to Barker's theory, intrauterine growth retardation is linked to the development of T2D and other diseases such as hypertension later in life (Barker, 2007). This risk is hypothesised to be as a result of epigenetic mechanisms programming the foetus to develop diseases such as T2D (Barker, 2007). There is a need for research into foetal programming among humans as there has been considerable evaluation of this phenomenon in animal studies but few among humans (Goh, 2012). Evidence from the Search for Diabetes in Youth Study indicated an association of T2D during pregnancy with early onset of the disease later in life (Dabelea & Pettitt, 2001). Protection from T2D risk up to the age of 21 years has been reported among breastfed infants (Mayer-Davis *et al.*, 2008). It is thus of critical importance to target T2D prevention initiatives in pregnant mothers.

2.3.3 Environmental chemicals

Environmental chemicals which cause the malfunctioning of hormones are termed endocrine-disrupting chemicals (EDCs) (Diamanti-Kandarakis *et al.*, 2009). The effects of these chemicals on T2D risk has been reported to be trans-generational, thereby suggesting that's these chemicals might cause changes to the genome or epigenome of individuals (Goh, 2012). A plasticiser like Bisphenol (BPA) has been reported to cause temporary hyperinsulinaemia and insulin resistance in the long term (Alonso-Magdalena *et al.*, 2006). In consequence, countries such as Canada have banned the use of BPA in the manufacturing of infant feeding bottles (Goh, 2012). Dioxin-related compounds found in petroleum-derived compounds and advanced glycation end products which are produced during cooking at high temperatures are other examples of EDCs which have been associated with T2D risk (Remillard & Bunce, 2002; Sandu *et al.*, 2005).

2.3.4 Microbiome

The gut microbiota has been reported to be associated with T2D risk (Goh, 2012). The microbiome contains 100 times more information than the human genome and secretes products that provide functions beyond the host genome, thereby making it pivotal to human physiology (Kahn *et al.*, 2014). Although the gut microbiota is postulated to be involved in the aetiology of T2D, the precise bacterial colonies involved have yet to be determined (Kahn *et al.*, 2014). Nonetheless, studies using faecal samples have noted metagenomic markers which are associated with T2D and which vary among ethnic groups (Karlsson *et al.*, 2013). The infusion of the intestinal microbiota of lean individuals among research participants with metabolic syndrome led to improved insulin sensitivity, thereby supporting the role of gut microbiota in T2D aetiology (Vrieze *et al.*, 2012).

2.4 GENETICS OF T2D

Evidence of the strong heritability of T2D exists, indicated by concordance rates of this disease in twin studies (Florez, 2013). Monozygotic twins possess similar genetic

material, and have concordance rates up to 96% for T2D (Newman *et al.*, 1987). Dizygotic twins, who are non-identical, were noted to have concordance rates of 20-30% (Medici *et al.*, 1999). A 40% risk of developing T2D has been noted for offspring of diabetics (Florez, 2013). The heritability estimates of T2D are estimated at 0.49 (Risch, 1990). However, the precise genetic factors that constitute the strong heritability component of T2D are yet to be fully clarified. Various methods have been applied to explore these genetic factors and the Genome Wide Association Study (GWAS) approach has led to the identification of many markers in comparison with prior techniques. Therefore, the coming discussion will address the identification of T2D genetic factors before and after the GWAS era.

2.4.1 Pre-GWAS era

Linkage analysis and candidate gene approaches were among the earlier methods used to dissect the heritability of T2D (Sanghera & Blackett, 2012). Linkage analysis involved the genotyping of 400-500 genetic markers across the genome, then evaluating the shared stretches of linkage disequilibrium (LD) among the affected and unaffected related individuals to determine disease-causing loci (Ahlqvist *et al.*, 2011). Through the linkage analysis approach, disease loci for Mendelian diseases were determined (Sanghera & Blackett, 2012). However, this approach did not yield many genetic loci for complex and common diseases like T2D. Of significance is the identification of calpain 10 (*CAPN10*) and transcription factor 7-like 2 (*TCF7L2*) genes which were also later identified through GWAS strategies, with the rs7093146 SNP being noted to have the highest effect size on T2D in European populations (Groop & Pociot, 2014).

The candidate gene approach changed focus from mapping LD stretches in family-based studies, as in the linkage analysis approach, to a population-based approach for determining disease-related loci (Groop & Pociot, 2014). The candidate genes were selected based on existing knowledge of these loci and their relation to the disease trait

(Ahlqvist *et al.*, 2011). However, this approach did not lead to the identification of many loci related to T2D owing to the limited knowledge of the full function of the genes (Groop & Pociot, 2014). The 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*) were the six candidate genes that consistently associated with T2D using this approach and these genes were later confirmed by GWASs (Ahlqvist *et al.*, 2011).

2.4.2 GWAS of T2D in European populations

The GWAS uses the principle of linkage analysis as noted in the earlier methods described above (Bush & Moore, 2012). However, it focuses on the most common markers of genetic variation in the human genome called the single nucleotide polymorphisms (SNPs) (Bush & Moore, 2012). This approach was made feasible as a result of the sequencing of the whole human genome and initiatives such as the HapMap project, which focused on cataloguing the genetic variation in humans (Sanghera & Blackett, 2012). It was noted that 500 000 to 1 000 000 tag SNPs could be used to account for more than 80% of the common variation in the human genome (Li *et al.*, 2008). Information regarding the tag SNPs was used to make genotyping arrays for assessing genetic variation in GWAS (Bush & Moore, 2012). The GWAS designs can be structured into case-control or quantitative study designs. In the case-control study design, the loci associated with a certain disease phenotype is evaluated by comparing its frequency and effects among patients (cases) and non-patients (controls) (Bush & Moore, 2012). In the quantitative trait design, the SNPs associated with a biomarker of a particular disease are assessed (Bush & Moore, 2012). Since tag SNPs are used in GWAS studies, the identified SNPs are not regarded as causal but in linkage disequilibrium with the causal variant (Bush & Moore, 2012). Thus fine-mapping strategies are required to find the causal variants (Bush & Moore, 2012).

The earliest GWASs were able to replicate the associations which had been depicted by the prior methods such as candidate gene and linkage analysis. These loci comprised of the 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*. From 2007-2008 a number of more loci continued to be discovered through GWASs by research groups of Zeggini *et al.*, (2007) and Scott *et al.*, (2007), as indicated in Table 2.2, including the discovery of the fat mass and obesity-associated (*FTO*) gene, which was noted to be associated with T2D through obesity (Scott *et al.*, 2007; Zeggini *et al.*, 2007). However, a new perspective came of combining the GWAS studies into meta-analysis to boost power to detect more SNPs associated with T2D.

Using 10 128 research participants and 2.2 million SNPs the Diabetes Genetics Replication and Meta-analysis (DAIGRAM) Consortium was developed and was able to depict six novel loci (Zeggini *et al.*, 2008). These loci consisted of 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 metalloproteinase with thrombospondin type 1 motif, 9 (*ADAMTS9*), and notch 2 (*NOTCH2*) (Zeggini *et al.*, 2008). A much bigger meta-analysis was initiated, termed the DIAGRAM plus consortium. In this consortium there were merges among the DIAGRAM consortium, the Diabetes Gene Discovery group, the KORA group, the Rotterdam study, the EUROSPAN and the deCODE genetics group, thereby lead to a total of 22 044 participants (Voight *et al.*, 2010). Twelve novel loci were then discovered in this initiative, illustrated in Table 2.2 (Voight *et al.*, 2010). Later on, among the European populations the Meta-analysis of Glucose and Insulin-related Traits Consortium (MAGIC) group was formed by combining the results of 21 GWASs (Dupuis *et al.*, 2010). Seventeen novel loci were identified which associated with glycaemic traits and only five of these SNPs were associated with T2D (Dupuis *et al.*, 2010).

2.4.3 GWAS of T2D among Asians

Although 96% of the GWASs were conducted among Europeans, additional loci were also detected among the Asians. The ubiquitin-conjugating enzyme E2E2 (*UBE2E2*) and eight more novel loci were identified through GWAS and GWAS meta-analysis among East Asians (Yasuda *et al.*, 2008; Cho *et al.*, 2011). Six novel loci were also discovered among South Asians (Kooner *et al.*, 2011). However, some of the earlier SNPs from the Europeans were not replicated in Asians. Conversely, there were also loci such as *KCNQ1*, which was discovered among the Japanese and later associated with T2D in Europeans; this locus was characterised as being involved in insulin resistance. (Sanghera & Blackett, 2012). Thus it was noted that GWASs in one population group were not adequate to determine the genetic architecture of T2D and for identifying variants applicable in varied populations, owing to the genetic diversity exhibited among different population groups (Sanghera & Blackett, 2012).

2.4.4 Trans-ethnic GWAS of T2D

GWASs of T2D in different ethnicities have been explored to determine the genetic markers of this disease. The LD differences among divergent population groups in transethnic meta-analysis of GWASs is pivotal for fine mapping, thereby allowing for the detection of causal variants and prioritisation of candidate genes (Li & Keating, 2014). However, the differences in allele frequencies and the directions of their effect need to be considered in such analyses. Methods and software such as the MANTRA, which allow for such heterogeneity have been developed (Morris, 2011). One of the largest trans-ethnic GWAS meta-analyses consisted of 26 488 cases and 83 964 controls selected from European, East Asian, South Asian, Mexican and Mexican American ancestry (Mahajan *et al.*, 2014). Seven novel loci were identified and directional consistency of T2D risk alleles were observed across the ancestries that were evaluated (Mahajan *et al.*, 2014).

2.4.5 GWAS and missing heritability

A total of about 135 SNPs associated with T2D, have been discovered thus far through GWASs (Prasad & Groop, 2015). Although much effort has been invested in boosting sample sizes in the conduction of large meta-analyses, the heritability explained by the SNPs associated with T2D is very low at 10% and a cause of concern (Voight *et al.*, 2010). Various perspectives that emanate from limitations of the GWAS approach are regarded as possible explanations to this missing heritability (Sanghera & Blackett, 2012). These factors consist of the failure to account for other forms of variation such as copy number variants and rare variants, gene-gene interactions, gene-environment interactions, epigenetics and a systems biology perspective (Sanghera & Blackett, 2012).

Table 2.2 List of common variants that have been discovered through GWAS

SNP	Chr	Pos	Genes	Context	Risk allele	Risk allele freq	P-value	OR	Reference
rs17106184	1	50,444,313	FAF1	Intron	G	0.92	4×10^{-9}	1.10	(Mahajan <i>et al.</i> , 2014)
rs10923931	1	119,975,336	NOTCH2, ADAM30	Intron	T	0.11	4×10^{-8}	1.13	(Zeggini <i>et al.</i> , 2008)
rs340874	1	213,985,913	PROX1	Intron	C	0.52	7×10^{-10}	1.07	(Dupuis <i>et al.</i> , 2010)
rs780094	2	27,518,370	GCKR	Intron	T	0.39	1×10^{-9}	1.06	(Dupuis <i>et al.</i> , 2010)
rs7578597	2	43,505,684	THADA	Missense	T	0.90	1×10^{-9}	1.15	(Zeggini <i>et al.</i> , 2008)
rs243021	2	60,357,684	BCL11A	Intergenic	A	0.50	3×10^{-15}	1.08	(Voight <i>et al.</i> , 2010)
rs7560163	2	150,781,422	RBM43, RND3	Intergenic	C	0.86	7×10^{-9}	1.33	(Palmer <i>et al.</i> , 2012)

rs7593730	2	160,314,943	ITGB6, RBMS1	Intron	C	0.78	4×10^{-8}	1.11	(Qi <i>et al.</i> , 2010)
rs3923113	2	164,645,339	GRB14	Intergen ic	A	0.74	1×10^{-8}	1.09	(Kooner <i>et al.</i> , 2011)
rs7578326	2	226,155,937	IRS1	Intron	A	0.71	5×10^{-20}	1.11	(Voight <i>et al.</i> , 2010)
rs1801282	3	12,351,626	PPARG	Missens e	C	0.88	6×10^{-10}	1.16	(Mahajan <i>et al.</i> , 2014)
rs7612463	3	23,294,959	UBE2E2	Intron	C	0.87	7×10^{-9}	1.10	(Mahajan <i>et al.</i> , 2014)
rs831571	3	64,062,621	PSMD6	Intergen ic	C	0.61	8×10^{-11}	1.09	(Cho <i>et al.</i> , 2011)
rs4607103	3	64,726,228	ADAMT S9	Intron	C	0.76	1×10^{-8}	1.09	(Zeggini <i>et al.</i> , 2008)
rs11717195	3	123,363,551	ADCY5	Intron	T	0.78	2×10^{-8}	1.09	(Mahajan <i>et al.</i> , 2014)
rs4402960	3	185,793,899	IGF2BP 2	Intron	T	0.29	2×10^{-9}	1.17	(Saxena <i>et al.</i> , 2007)
rs16861329	3	186,948,673	ST6GAL 1	Intron	G	0.75	3×10^{-8}	1.09	(Kooner <i>et al.</i> , 2011)
rs6808574	3	188,022,735	LPP	Intergen ic	C	0.79	6×10^{-9}	1.07	(Mahajan <i>et al.</i> , 2014)
rs6815464	4	1,316,113	MAEA	Intron	C	0.58	2×10^{-20}	1.13	(Cho <i>et al.</i> , 2011)
rs1801214	4	6,301,295	WFS1	Missens e	T	0.73	3×10^{-8}	1.13	(Voight <i>et al.</i> , 2010)
rs6813195	4	152,599,323	TMEM1 54	Intergen ic	C	0.59	4×10^{-14}	1.08	(Mahajan <i>et al.</i> , 2014)
rs702634	5	53,975,590	ARL15	Intron	A	0.76	7×10^{-9}	1.06	(Mahajan <i>et al.</i> , 2014)
rs4457053	5	77,129,124	ZBED3	Intron	G	0.20	3×10^{-12}	1.08	(Voight <i>et al.</i> , 2010)

rs9502570	6	7,258,384	RREB1, SSR1	Intergen ic	A	0.45	1×10^{-9}	1.06	(Mahajan <i>et al.</i> , 2014)
rs7754840	6	20,661,019	CDKAL 1	Intron	C	0.36	4×10^{-11}	1.12	(Scott <i>et al.</i> , 2007)
rs3132524	6	31,168,937	TCF19, POU5F 1	Intron	G	0.38	4×10^{-9}	1.07	(Mahajan <i>et al.</i> , 2014)
rs9470794	6	38,139,068	ZFAND 3	Intron	C	0.27	2×10^{-10}	1.12	(Cho <i>et al.</i> , 2011)
rs1535500	6	39,316,274	KCNK1 6	Missens e	T	0.42	2×10^{-8}	1.08	(Cho <i>et al.</i> , 2011)
rs1048886	6	70,579,486	C6orf57	Missens e	G	0.18	3×10^{-8}	1.54	(Sim <i>et al.</i> , 2011)
Rs2191349	7	15,024,684	DGKB, TMEM1 95	Intergen ic	T	0.52	1×10^{-8}	1.06	(Dupuis <i>et al.</i> , 2010)
rs864745	7	28,140,937	JAZF1	Intron	T	0.50	5×10^{-14}	1.10	(Zeggini <i>et al.</i> , 2008)
rs4607517	7	44,196,069	GCK	Intergen ic	A	0.16	5×10^{-8}	1.07	(Dupuis <i>et al.</i> , 2010)
rs6467136	7	127,524,904	PAX4, GCC1	Intergen ic	G	0.79	5×10^{-11}	1.11	(Cho <i>et al.</i> , 2011)
rs791595	7	128,222,749	MIR129, LEP	Intron	A	0.08	3×10^{-13}	1.17	(Hara <i>et al.</i> , 2014)
rs972283	7	130,782,095	KLF14	Intergen ic	G	0.69	2×10^{-10}	1.07	(Voight <i>et al.</i> , 2010)
rs515071	8	41,661,944	ANK1	Missens e	G	0.79	1×10^{-8}	1.18	(Imamura <i>et al.</i> , 2012)
rs896854	8	94,948,283	TP53IN P1	Missens e	T	0.24	1×10^{-9}	1.06	(Voight <i>et al.</i> , 2010)
rs13266634	8	117,172,544	SLC30A 8	Missens e	C	0.69	5×10^{-8}	1.12	(Zeggini <i>et al.</i> , 2007)

rs7041847	9	4,287,466	GLIS3	Intron	A	0.41	2×10^{-14}	1.10	(Cho <i>et al.</i> , 2011)
rs17584499	9	8,879,118	PTPRD	Intron	T	0.06	9×10^{-10}	1.57	(Tsai <i>et al.</i> , 2010)
rs10811661	9	22,134,095	CDKN2A, CDKN2B	Intergenic	T	0.83	5×10^{-8}	1.20	(Saxena <i>et al.</i> , 2007)
rs13292136	9	79,337,213	CHCHD9	Intergenic	C	0.90	3×10^{-8}	1.11	(Voight <i>et al.</i> , 2010)
rs11787792	9	136,357,696	GPSM1	Intron	A	0.87	2×10^{-10}	1.15	(Hara <i>et al.</i> , 2014)
rs12779790	10	12,286,011	CDC123, CAMK1D	Intergenic	G	0.18	1×10^{-10}	1.11	(Zeggini <i>et al.</i> , 2008)
rs1802295	10	69,171,718	VPS26A	UTR-3	A	0.26	4×10^{-8}	1.08	(Kooner <i>et al.</i> , 2011)
rs12571751	10	79,182,874	ZMIZ1	Intron	A	0.51	2×10^{-10}	1.09	(Mahajan <i>et al.</i> , 2014)
rs1111875	10	92,703,125	HHEX	Intergenic	C	0.52	6×10^{-10}	1.13	(Scott <i>et al.</i> , 2007)
rs7903146	10	112,998,590	TCF7L2	Intron	T	0.30	2×10^{-34}	1.65	(Sladek <i>et al.</i> , 2007)
rs10886471	10	119,389,891	GRK5	Intron	C	0.78	7×10^{-9}	1.12	(Li <i>et al.</i> , 2013)
rs3842770	11	2,157,440	INS-IGF2	Intron	A	0.23	3×10^{-8}	1.14	(Ng, 2015)
rs2237892	11	2,818,521	KCNQ1	Intron	C	0.61	2×10^{-42}	1.40	(Yasuda <i>et al.</i> , 2008)
rs5215	11	17,387,083	KCNJ11	Missense	C	0.27	5×10^{-11}	1.14	(Zeggini <i>et al.</i> , 2007)
rs1552224	11	72,722,053	CENTD2	Intron	A	0.45	1×10^{-22}	1.14	(Voight <i>et al.</i> , 2010)
rs1387153	11	92,940,662	MTNR1B	Intergenic	T	0.35	8×10^{-15}	1.09	(Voight <i>et al.</i> , 2010)

rs1531343	12	65,781,114	HMG2A	Intron	C	0.21	4×10^{-9}	1.10	(Voight <i>et al.</i> , 2010)
rs7961581	12	71,269,322	LGR5, TSPAN8	Intergenic	C	0.27	1×10^{-9}	1.09	(Zeggini <i>et al.</i> , 2008)
rs7957197	12	121,022,883	HNF1A	Intron	T	0.89	2×10^{-8}	1.07	(Voight <i>et al.</i> , 2010)
rs1727313	12	123,156,306	MPHOSPH9	NcRNA	C	0.50	1×10^{-8}	1.06	(Mahajan <i>et al.</i> , 2014)
rs9552911	13	23,290,518	SGCG, SACS	Intron	G	0.93	2×10^{-8}	1.49	(Saxena <i>et al.</i> , 2007)
rs1359790	13	80,143,021	SPRY2	Intergenic	G	0.71	6×10^{-9}	1.15	(Shu <i>et al.</i> , 2010)
rs7403531	15	38,530,704	RASGRP1	Intron	T	0.35	4×10^{-9}	1.10	(Li <i>et al.</i> , 2013)
rs7172432	15	62,104,190	C2CD4B, C2CD4A	Intergenic	A	0.58	9×10^{-14}	1.11	(Yamauchi <i>et al.</i> , 2010)
rs7178572	15	77,454,848	HMG20A	Intron	G	0.70	1×10^{-8}	1.11	(Perry <i>et al.</i> , 2012)
rs11634397	15	80,139,880	ZFAND6	Intergenic	G	0.44	2×10^{-9}	1.06	(Voight <i>et al.</i> , 2010)
rs2028299	15	89,831,025	AP3S2	NcRNA	C	0.31	2×10^{-11}	1.10	(Kooner <i>et al.</i> , 2011)
rs8042680	15	90,978,107	PRC1	Intron; intron	A	0.74	2×10^{-10}	1.07	(Voight <i>et al.</i> , 2010)
rs8050136	16	53,782,363	FTO	Intron	A	0.38	1×10^{-12}	1.17	(Scott <i>et al.</i> , 2007)
rs391300	17	2,312,964	SRR	Intron	G	0.62	3×10^{-9}	1.28	(Tsai <i>et al.</i> , 2010)
rs312457	17	7,037,074	SLC16A13	Intron	G	0.08	8×10^{-13}	1.20	(Hara <i>et al.</i> , 2014)
rs4430796	17	37,738,049	HNF1B	Intron	G	0.28	2×10^{-11}	1.19	(Li <i>et al.</i> , 2013)

rs8090011	18	7,068,463	LAMA1	Intron	G	0.38	8×10^{-9}	1.13	(Perry <i>et al.</i> , 2012)
rs12970134	18	60,217,517	MC4R	Intergenic	A	0.27	3×10^{-8}	1.08	(Mahajan <i>et al.</i> , 2014)
rs3786897	19	33,402,102	PEPD	Intron	A	0.56	1×10^{-8}	1.10	(Cho <i>et al.</i> , 2011)
rs4812829	20	44,360,627	HNF4A	Intron	A	0.29	3×10^{-10}	1.09	(Kooner <i>et al.</i> , 2011)

Chr = chromosome; Pos = position; A = adenine; G = guanine; C = cytosine; T = thymine; OR = Odds ratio

The GWAS approach focuses mainly on common variants which have allele frequencies greater than 5% (Bush & Moore, 2012). It has been postulated that rare variants might have larger effect sizes for disease risk and these are usually omitted in GWASs (Sanghera & Blackett, 2012). However, a recent large-scale study among 44,414 individuals of European ancestry has challenged this notion by indicating that the low and rare variants might not have the large effects that they are postulated to have (Fuchsberger *et al.*, 2016). Although this study comprised of participants from South Asia, East Asia, Hispanics and African Americans, the role of rare variant associations with T2D among African ethnicities are yet to be explored. (Fuchsberger *et al.*, 2016). The advent of next generation sequencing and the reduction in costs of sequencing offer an opportunity to generate more African whole genomes which can be used to explore the contribution of rare variants to T2D risk among African population groups. Through extreme phenotype sampling, smaller sample sizes than those required for random sampling, can be used to interrogate the associations of rare variants with T2D risk using whole genome sequence data. However, to the best of our knowledge, such an initiative is yet to be explored in Africa.

2.4.6 Extreme phenotype sampling and WGS

Extreme phenotype sampling involves the sampling of people with extreme values of a quantitative trait. This approach was known historically as “Selective genotyping” and it was first described by Lander and Botstein in animal studies (Lander & Botstein, 1989). Later, Carey *et al.* (1991) used this approach in linkage analysis and noted that it boosted power to detect loci of quantitative traits (Carey & Williamson, 1991). It is believed that complex diseases such as T2D represent extremes of quantitative traits hence the renewed interest in applying extreme phenotype sampling in whole-genome sequencing association studies (Guey *et al.*, 2011; Fuchsberger *et al.*, 2016). Typically, a quantitative trait is assumed to follow a normal distribution in which the samples are selected from the largest and smallest percentiles. However, for disease traits, known risk factors such as early onset and family history can be used to select extremes (Guey *et al.*, 2011). The unexplained variation of quantitative traits besides those of known covariates is deemed to be as a result of genetic factors. The regression residuals computed after adjusting for known risk factors indicate the unexplained variation of the trait, and are used for extreme phenotype sampling (Lanktree *et al.*, 2010). Studies of the extremes of unexplained variation have been noted to enable selection of participants whose traits are explained rather by genetic factors, while omitting those explained by traditional risk factors (Lanktree *et al.*, 2010).

Extreme phenotype sampling has been reported to boost the power of detecting rare variants as opposed to random sampling (Guey *et al.*, 2011). The nonsense variant in *SLC30A8* was discovered through this approach and later seen to be associated with 53% reduced T2D risk among 44 000 cases and controls (Flannick *et al.*, 2014). However, methodological concerns have been raised about the analysis of rare variants in the extreme phenotype designs (Barnett *et al.*, 2013). Traditional rare-variant association tests require the dichotomising of the extreme traits (Guey *et al.*, 2011). However, such an approach leads to loss of power by decreasing the phenotype information (Barnett *et al.*, 2013). Recent methods which treat the extreme phenotypes as continuous phenotypes, while also using the optimal rare-variant association test, the Sequence

Kernel Association Test (SKAT-O), have been explored through simulation studies and improved rare variant detection (Barnett *et al.*, 2013). More studies are required to validate this approach.

2.4.7 Polygenic risk scores and T2D

T2D is postulated to have a polygenic aetiology (Fuchsberger *et al.*, 2016). However, such a phenomenon, which indicates the interaction of SNPs, is not accounted for in GWAS methodologies (Sanghera & Blackett, 2012). The additive genetic model used in the GWAS analysis assumes that all the SNPs have equal effects (Groop & Pociot, 2014). However, these common variants have been reported to interact, resulting in a phenotypic outcome that is different from the additive effects of the individual loci, a concept called epistasis (Churchill, 2013). It remains a challenge to incorporate the concept of epistasis in GWAS methodologies although some probable approaches by Evans *et al.* (2006) have been recommended and more research is still required in this area to unravel the missing heritability (Groop & Pociot, 2014).

On the other hand, there are proponents of the view that we have reached the limit of discovering more SNPs associated with T2D and that we should now shift focus to developing polygenic risk scores to account for the missing heritability of T2D (Vassy *et al.*, 2014). Although the concept of epistasis is not featured in these polygenic risk scores, efforts have been made to capture the varied effects of these SNPs on phenotypes by weighting them by their effect sizes in the generation of the polygenic risk scores (Smith *et al.*, 2015). However, the effect sizes should be selected from meta-analysis to reduce the bias as result varied study specific effect sizes of the SNPs used to construct the polygenic risk scores (Smith *et al.*, 2015). Polygenic risks scores have also been developed based on the pathways in which the characterised SNPs are involved in, so as to determine polygenic scores that are strongly associated with T2D risk (Vassy *et al.*, 2014). However, it has been noted that the polygenic risk scores, modestly improve the

prediction of T2D risk when combined with to known risk factors in Europeans (Wang *et al.*, 2015). The predictive utility of the polygenic risk scores associated with T2D among African populations has yet to be determined. A study by Vassy *et al.* (2014) indicated that, regardless of the genetic differences and LD patterns among Caucasians and African Americans, similar polygenic risks were associated with T2D (Vassy *et al.*, 2014).. Therefore, the results of Vassy *et al.* (2014) suggest that the GRS is not limited by genetic differences among varied ethnicities, and this phenomenon needs to be explored further in other African ancestries (Vassy *et al.*, 2014).

2.4.8 GWAS of T2D in people of Africa Ancestry

Ninety percent of GWAS studies for T2D have been conducted among Europeans (Sanghera & Blackett, 2012). A number of studies were done in the African Americans regardless that the genotyping arrays had less coverage of African variants. Three established loci of TCF7L2, KCNQ1 and HMGA2 were replicated among the African Americans (Ng, 2015). Nonetheless, novel variants in the HLA-B and INS-IGF2 were noted to be more frequent in the African Americans and associated with T2D compared to the Caucasians. Such a phenomenon had been reported with regards to UBE2E2 which was noted to be more common in East Asians and associated with T2D while being rare or monomorphic in Caucasians (Ng, 2015). Thereby suggesting the existence of private and population specific variants to be associated with T2D risk.

T2D is projected to increase most in low-income countries in Africa which are ill-equipped to manage this epidemic. To date, there is not a single GWAS which has been performed for T2D on the African continent (Ramsay, 2012; Ramsay *et al.*, 2016). However, some studies are now underway in the Human Hereditary and Health initiative to determine the genetic and environmental determinants of T2D. Genetic research in Africa has been hampered by lack of resources however the joint funding from the NIH and the Wellcome Trust have helped to kick-start extensive genetic research in Africa (Ramsay *et al.*, 2016).

Recently a systematic review and meta-analysis by Yako *et al.* (2016) reviewed 60 studies that have been conducted to evaluate the genetics of T2D in Africa (Yako *et al.*, 2016). Most of the studies were replication studies of variants that had been associated with T2D in Europeans and Asians, besides the genome-wide scans that were done in West Africa (Yako *et al.*, 2016). The sample sizes were modest, ranging from 30 to 1052, and the majority of the studies were done in North Africa (Yako *et al.*, 2016). There were no novel SNPs documented in these studies and *TCF7L2* was the loci which associated with T2D among most African ethnicities (Yako *et al.*, 2016). Therefore, scant genetic information, and that mostly from North Africa, is available and more genetic studies are required to unravel the genetic architecture of T2D among Africans.

2.4.9 Genetic diversity among Africans

Africans residing in Africa have been reported to harbour the greatest genetic diversity in the world (Ramsay, 2012). The differences in the genetic diversity of Africans, Caucasians and Asians have also been depicted with regard to LD, which indicates how genetic variants are related (Teo *et al.*, 2010). It has been observed that African populations have shorter LD blocks while the Caucasians and Asians have longer LD blocks (Teo *et al.*, 2010). Added to this, differences, in allele frequencies have been noted to exist among different populations thus making it difficult to replicate the discovery of variants from Caucasian and Asians in African populations (Teo *et al.*, 2010). The genotyping arrays have also been reported to capture less of the African genetic variation (May *et al.*, 2013b). However, an African customised genotyping chip has been developed with the use of the limited whole genome sequences in Africa. This chip will be used to determine the unique common genetic variation associated with diseases in Africa (Ramsay *et al.*, 2016).

2.4.10 Genetics of T2D in South Africa

Limited studies exist evaluating the genetics of T2D among black South Africans (Chikowore *et al.*, 2015). A study by Pirie *et al.* (2010) among the Zulus noted the association of *TCF7L2* with T2D (Pirie *et al.*, 2010). However, another study by Chikowore *et al.* (2015) was not able to replicate the association of *TCF7L2* with T2D risk and replicated signals in the *FTO*, *PSMD6*, *C2CD4B*, *SLC44A3* and *F3* genes among the Setswana black South African population (Chikowore *et al.*, 2015). The discrepancies in the findings of these studies might be as a result of the genetic differences between these ethnolinguistic groups and or due to the modest sample sizes which might have been less powered to replicate the evaluated SNPs. Nonetheless, more studies using varied methodologies are required to unravel the genetic determinants of T2D among black South Africans. These strategies include GWAS meta-analyses, genotype imputation and use of whole genome sequence data for association analysis. Using the whole genome sequence data it will also be pivotal to determine the how the nine ethnolinguistic ethnicities among the black South Africans are related.

Recently a number of whole genomes from the Zulu and the Sotho ethnolinguistic groups were sequenced through the African Human Genome Project (AGVP) (Gurdasani *et al.*, 2015). These, generated sequences can be used to assesses the genetic relatedness of the black South African ethnolinguistic. Great genetic diversity has been reported to exist among African ethnic groups (May *et al.*, 2013a). However, it might also be possible that some ethnolinguistic groups might share similar ancestries and have genetic similarities that will enable disease variants to be generalised among these population groups. Thus, efforts to determine the relatedness of the ethnolinguistic using whole genome sequence data, will help advance the unravelling of the genetic determinants of T2D among the black South Africans as the variants determined in one ethnolinguistic group might also be relevant in a related ethnolinguistic group.

The genetic determinants have been reported to be stable, however the surge in the T2D prevalence in T2D is postulated to be as a result of the recent adoption of dietary patterns and lifestyles which increase T2D risk by people with prior genetic predisposition to this

disease(Franks *et al.*, 2013). The dietary determinants are of T2D are elaborated further in the following sections.

2.5 NUTRITION AND T2D

2.5.1 Nutrition transition

The nutrition transition from traditional to Western eating patterns has been reported in a number of studies to be exacerbating the prevalence of chronic diseases which includes T2D (MacIntyre *et al.*, 2002; Vorster *et al.*, 2011). This phenomenon has been made possible by the changes in the food system in urban areas. A lot of supermarkets and fast foods restaurants have mushroomed in urban areas, making highly processed foods, sugary beverages and high-energy snacks easily accessible to many (Popkin *et al.*, 2012). Reduced physical activity in the context of the intake of high energy-dense foods has caused the levels of obesity to rise, making more people susceptible to developing T2D (Ley *et al.*, 2014). However, most nutrition transition dietary associations with non-communicable diseases (NCDs) like T2D in Sub Saharan Africa are from cross sectional studies which are prone to bias and need to be validated in longitudinal studies such as the Prospective Urban and Rural Epidemiological (PURE) study (Mbanya *et al.*, 2010). Limited evidence exists of the precise dietary components that are associated with increased T2D risk, as the nutrition transition that consists of increases in animal protein intake, fat and sugar is generalised for increasing disease risk of a number of lifestyle related NCDs such as hypertension. A review by Steyn and Temple. (2012) failed to find local evidence of the direct association of T2D with consumption of sugar (Steyn & Temple, 2012). Conversely, studies in developed countries were able to report that the consumption of sugar sweetened beverages was associated with increased T2D risk (Hu, 2013). Therefore, more local studies which assess dietary factors associated with T2D are required.

2.5.2 Macronutrients and T2D

The quantity and quality, of carbohydrates and fat associated with T2D risk has been a source of debate (Ley *et al.*, 2014). It had previously been postulated that high total fat intake leads to insulin resistance by indirectly promoting weight gain which will lead to insulin insensitivity (Riserus *et al.*, 2009). However, total fat intake was not associated with diabetes in a number of longitudinal studies. For instance, the incidence of T2D was not reduced in the Women's Health Initiative study, in the low-fat diet group compared with the control group (Tinker *et al.*, 2008). However, the quality of fat has been noted to be pivotal with regard to T2D risk. Diets that contain more plant-based oils over animal based fats were observed to be beneficial for lowering T2D risk (Hu *et al.*, 2001). In line with this, replacing saturated fat with omega -6 polyunsaturated fat was associated with lower T2D risk in the Nurses' Health Study (Salmeron *et al.*, 2001).

Evidence from longitudinal studies indicates that the quantity of carbohydrate in the diet has less effect on T2D risk than the quality of the carbohydrate (Hauner *et al.*, 2012). A meta-analysis of cohort studies indicated that dietary fibre from cereal products was associated with reduced T2D risk. Low glycaemic index (GI) and low glycaemic load (GL) have also been observed to be associated with reduced T2D risk in meta-analysis of longitudinal studies (Bhupathiraju *et al.*, 2014).

2.5.3 Vitamins and minerals and T2D

The associations of a number of micronutrients with T2D risk have been evaluated. Magnesium intake was inversely associated with T2D risk, and this effect was even noted among overweight individuals (Dong *et al.*, 2011). High heme-iron intake was associated with higher T2D risk (Ley *et al.*, 2014). Vitamin D has been associated with reduced T2D risk among diverse populations (Zhao *et al.*, 2012). However, Vitamin D was not able to improve fasting glucose, glycated haemoglobin or insulin sensitivity in randomised

controlled trials with modest sample sizes. More well powered studies will help unravel the role of Vitamin D with regard to T2D prevention (Mitri *et al.*, 2011).

2.5.4 Individual foods and T2D

The associations of particular foods and food groups with T2D risk have been assessed in prospective cohort studies. Consistent associations of whole grains, green leafy vegetables, nuts and coffee have been found with reduced T2D risk (Ley *et al.*, 2014). Conversely, sugar-sweetened beverages, red meats and processed meats like bacon, sausages and hot dogs have been associated with high risk of T2D (Ley *et al.*, 2014). The association of sea food with T2D varies with regional location. Increased fish consumption was associated with higher T2D risk in North America and Europe but lower T2D risk in Asia (Wu *et al.*, 2012). The cultural differences in food preparation are postulated to explain the varied regional associations of fish intake with T2D risk (Ley *et al.*, 2014).

2.5.5 Dietary patterns and T2D

Dietary pattern approaches assess the diet as a whole and are classified into *a priori* score based approaches and *a posteriori* data driven techniques such as the factor analysis technique (McEvoy *et al.*, 2014). The Prudent, Mediterranean and Vegetarian dietary patterns have been reported to lower T2D risk while the Western dietary patterns increase T2D risk. The Dietary Guidelines Alternative Healthy Eating Index (AHEI) is an example of a score based dietary pattern which has been associated with low T2D risk (Ley *et al.*, 2014). The components of the above mentioned dietary patterns are indicated in Table 2. However, the *a priori* score-based methods were noted to be of limited utility in evaluating the role of diet in CVD risk in the black South African population (Dolman *et al.*, 2014). Unlike the score based approach which computes the data into predefined dietary patterns using prior knowledge of diet and health, the data driven methods such as factor analysis, allow the data at hand to determine the dietary pattern (McEvoy *et al.*, 2014). This allows a more realistic dietary pattern to be derived which reflects the variety

and uniqueness that is a function of the different sociocultural norms that determine eating patterns (McEvoy *et al.*, 2014). However, limited evidence exists on the use of *a posteriori* techniques such as factor analysis to assess the role of dietary patterns in relation to T2D risk in non-Western populations like the South African population.

Table 2.3 Dietary patterns and their associations with T2D risk

Dietary Pattern	Food Components	Associations with T2D risk
Mediterranean diet	high consumption of minimally processed plant-based foods; olive oil as the principal source of fat; low- to-moderate consumption of dairy products, fish, and poultry; low consumption of red meat; and low-to- moderate consumption of wine with meals	Mediterranean dietary patterns were associated with lower risk of T2D in prospective cohort studies and RCTs.
Dietary Approaches to Stop Hypertension (DASH)	rich in vegetables, fruits, and low-fat dairy products, including whole grains, poultry, fish, and nuts; lower in saturated fat, red meat, sweets, and sugar containing beverages; and often reduced in sodium	Adherence to the DASH diet was associated with lower risk of T2D.
Vegetarian and vegan	vegan diets avoid all animal-derived products; vegetarian diets are devoid of some animal products and include lacto-ovo (consuming dairy and/or eggs), pesco (consuming fish, eggs, and/or dairy) and semi- vegetarian (consuming all but no red meat and poultry)	Vegan, lacto-ovo and semi-vegetarian diets were associated with lower risk of T2D.
Dietary guidelines - Alternative Healthy Eating Index (AHEI)	indices of the diet quality created based on foods and nutrients predictive of chronic disease risk, including greater intake of vegetables and fruits, whole grains, nuts and legumes, long-chain omega-3 fatty acids, PUFAs; lower intake of sugar-sweetened beverages and fruit juice, red/processed meat, trans fat, sodium; and moderate alcohol consumption	Adherence to high quality diet assessed by AHEI was strongly associated with lower risk of T2D.

Prudent pattern	dietary patterns higher in fruits, vegetables, whole grains, and vegetable fats and lower in red meats, refined grains, and sugared soft drinks	Prudent dietary patterns over Western dietary patterns were associated with lower T2D risk.
-----------------	---	---

Adopted from (Ley *et al.*, 2014)

2.5.6 Dietary patterns and T2D risk in black South Africans

Limited evidence exists regarding the association of dietary patterns with T2D. There is a need for dietary patterns approaches which assess the effects of the diet as a whole on T2D risk. Although, rapid urbanisation has led to the nutrition transition that is attributed to the surge in prevalence of NCDs such as T2D, it has been observed to lead to improvements in micronutrient intake, as a result of the improvements in fruit and vegetable intake (Hattingh *et al.*, 2008). This makes the association of dietary factors with T2D even more complex due to the coexistence of the intake of fruits and vegetables which protect against T2D risk and the intake of animal protein, sugar and fat intake which increase the risk of this disease, among urban residents in South Africa. It is therefore pivotal to evaluate the interaction of these dietary factors and their effect on T2D risk in the black South African population.

2.5.7 Nutrient patterns vs Dietary Patterns

Little information exists regarding the associations of nutrient patterns with T2D. Dietary patterns have been much explored and they are easy to translate into public health recommendations (Willett, 1998). However, as opposed to dietary patterns which are limited for comparison among population groups because of variation in foods consumed as a result of cultural norms, nutrients are universal and easy to compare among varied ethnic groups (Moskal *et al.*, 2014). Nutrients are more directly involved than foods in the metabolism and aetiology of diseases (Moskal *et al.*, 2014). This information might aid in understanding the aetiology of T2D and thus help in the development of preventative and

curative measures to curb this burden. Thus more studies of the associations of nutrient patterns with T2D risk are required.

2.6 SUMMARY

T2D is a complex disease that has an aetiology which has not been fully elucidated. However, the increase in the prevalence of T2D is postulated to be a result of the adoption of Westernised lifestyles among people who are genetically and epigenetically predisposed to develop this disease. However, it is not yet clear which dietary and genetic factors contribute to T2D risk in black South Africans. Thus, more studies are required to determine the genetic and dietary determinants of T2D in black South Africans. In this study we sought to evaluate the genetic factors in the form of SNPs associated with T2D risk using polygenic risk scores and whole genome sequences. This aspects together with the evaluation of nutrient patterns associated with glycated hemoglobin and fasting glucose are elaborated in the chapters that follow.

2.7 REFERENCES

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.

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

American Diabetes, A. 2013. Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 36 Suppl 1:S67-74.

Amod, A., Ascott-Evans, B.H., Berg, G.I., Blom, D.J., Brown, S.L., Carrihill, M.M., Dave, J.A., Distiller, L.A., Ganie, Y.N., Grobler, N., Heilbrunn, A.G., Huddle, K.R.L., Janse van Rensburg, G., Jivan, D., Joshi, P., Khutsoane, D.T., Levitt, N.S., May, W.M., Mollentze, W.F., Motala, A.A., Paruk, I.M., Pirie, F.J., Raal, F.J., Rauuff, S., Raubenheimer, P.J., Randeree, H.A.R., Rheeder, P., Tudhope, L., Van Zyl, D.J. & M., Y. 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.

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

Barnett, I.J., Lee, S. & Lin, X. 2013. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet Epidemiol*, 37(2):142-151.

Bhupathiraju, S.N., Tobias, D.K., Malik, V.S., Pan, A., Hruby, A., Manson, J.E., Willett, W.C. & Hu, F.B. 2014. Glycemic index, glycemic load, and risk of type 2 diabetes: results from 3 large US cohorts and an updated meta-analysis. *Am J Clin Nutr*, 100(1):218-232.

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

Carey, G. & Williamson, J. 1991. Linkage analysis of quantitative traits: increased power by using selected samples. *Am J Hum Genet*, 49(4):786-796.

Chikowore, T., Conradie, K.R., Towers, G.W. & van Zyl, T. 2015. Common Variants Associated with Type 2 Diabetes in a Black South African Population of Setswana Descent: African Populations Diverge. *OMICS*, 19(10):617-626.

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., Consortium, D., Mu, T.C., 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. *Nat Genet*, 44(1):67-72.

Churchill, G.A. 2013. Epistasis. (In Maloy, S. & Hughes, K., eds. *Brenner's Encyclopedia of Genetics* (Second Edition). San Diego: Academic Press. p. 505-507).

Cnop, M., Landchild, M.J., Vidal, J., Havel, P.J., Knowles, N.G., Carr, D.R., Wang, F., Hull, R.L., Boyko, E.J., Retzlaff, B.M., Walden, C.E., Knopp, R.H. & Kahn, S.E. 2002. The concurrent accumulation of intra-abdominal and subcutaneous fat explains the association between insulin resistance and plasma leptin concentrations : distinct metabolic effects of two fat compartments. *Diabetes*, 51(4):1005-1015.

Cnop, M., Vidal, J., Hull, R.L., Utzschneider, K.M., Carr, D.B., Schraw, T., Scherer, P.E., Boyko, E.J., Fujimoto, W.Y. & Kahn, S.E. 2007. Progressive loss of beta-cell function leads to worsening glucose tolerance in first-degree relatives of subjects with type 2 diabetes. *Diabetes Care*, 30(3):677-682.

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.

Diamanti-Kandarakis, E., Bourguignon, J.P., 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.

Dolman, R.C., Wentzel-Viljoen, E., Jerling, J.C., Feskens, E.J., Kruger, A. & Pieters, M. 2014. The use of predefined diet quality scores in the context of CVD risk during urbanization in the South African Prospective Urban and Rural Epidemiological (PURE) study. *Public Health Nutrition*, 17(8):1706-1716.

Dong, J.Y., Xun, P.C., He, K. & Qin, L.Q. 2011. Magnesium Intake and Risk of Type 2 Diabetes Meta-analysis of prospective cohort studies. *Diabetes Care*, 34(9):2116-2122.

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., Roccascaccia, 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., Gyllenstein, 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., Lecoœur, 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., Mwitteman, 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., Ruukonen, A., RiiitaJarvelin, 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. *Nat Genet*, 42(2):105-116.

Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B., Grarup, N., Burt, N.P., Mahajan, A., Fuchsberger, C., Atzmon, G., Benediktsson, R., Blangero, J., Bowden, D.W., Brandslund, I., Brosnan, J., Burslem, F., Chambers, J., Cho, Y.S., Christensen, C., Douglas, D.A., Duggirala, R., Dymek, Z., Farjoun, Y., Fennell, T., Fontanillas, P., Forsen, T., Gabriel, S., Glaser, B., Gudbjartsson, D.F., Hanis, C., Hansen, T., Hreidarsson, A.B., Hveem, K., Ingelsson, E., Isomaa, B., Johansson, S., Jorgensen, T., Jorgensen, M.E., Kathiresan, S., Kong, A., Kooner, J., Kravic, J., Laakso, M., Lee, J.Y., Lind, L., Lindgren, C.M., Linneberg, A., Masson, G., Meitinger, T., Mohlke, K.L., Molven, A., Morris, A.P., Potluri, S., Rauramaa, R., Ribel-Madsen, R., Richard, A.M., Rolph, T., Salomaa, V., Segre, A.V., Skarstrand, H., Steinthorsdottir, V., Stringham, H.M., Sulem, P., Tai, E.S., Teo, Y.Y., Teslovich, T., Thorsteinsdottir, U., Trimmer, J.K., Tuomi, T., Tuomilehto, J., Vaziri-Sani, F., Voight, B.F., Wilson, J.G., Boehnke, M., McCarthy, M.I., Njolstad, P.R., Pedersen, O., Go, T.D.C., Consortium, T.D.G., Groop, L., Cox, D.R., Stefansson, K. & Altshuler, D. 2014. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet*, 46(4):357-363.

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., 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.

Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., Rivas, M.A., Perry, J.R.B., Sim, X., Blackwell, T.W., Robertson, N.R., Rayner, N.W., Cingolani, P., Locke, A.E., Tajas, J.F., Highland, H.M., Dupuis, J., Chines, P.S., Lindgren, C.M., Hartl, C., Jackson, A.U., Chen, H., Huyghe, J.R., van de Bunt, M., Pearson, R.D., Kumar, A., Muller-Nurasyid, M., Grarup, N., Stringham, H.M., Gamazon, E.R., Lee, J., Chen, Y.H., Scott, R.A., Below, J.E., Chen, P., Huang, J., Go, M.J., Stitzel, M.L., Pasko, D., Parker, S.C.J., Varga, T.V., Green, T., Beer, N.L., Day-Williams, A.G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M.K., Kim, B.J., Kim, Y., Kim, Y.J., Kwon, M.S., Lee, J., Lee, S., Lin,

K.H., Maxwell, T.J., Nagai, Y., Wang, X., Welch, R.P., Yoon, J., Zhang, W., Barzilai, N., Voight, B.F., Han, B.G., Jenkinson, C.P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M.C.Y., Palmer, N.D., Balkau, B., Stancakova, A., Abboud, H.E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J.D., Neale, B.M., Purcell, S., Butterworth, A.S., Howson, J.M.M., Lee, H.M., Lu, Y.C., Kwak, S.H., Zhao, W., Danesh, J., Lam, V.K.L., Park, K.S., Saleheen, D., So, W.Y., Tam, C.H.T., Afzal, U., Aguilar, D., Arya, R., Aung, T., Chan, E., Navarro, C., Cheng, C.Y., Palli, D., Correa, A., Curran, J.E., Rybin, D., Farook, V.S., Fowler, S.P., Freedman, B.I., Griswold, M., Hale, D.E., Hicks, P.J., Khor, C.C., Kumar, S., Lehne, B., Thuillier, D., Lim, W.Y., Liu, J., van der Schouw, Y.T., Loh, M., Musani, S.K., Puppala, S., Scott, W.R., Yengo, L., Tan, S.T., Taylor, H.A., Thameem, F., Wilson, G., Wong, T.Y., Njolstad, P.R., Levy, J.C., Mangino, M., Bonnycastle, L.L., Schwarzmayr, T., Fadista, J., Surdulescu, G.L., Herder, C., Groves, C.J., Wieland, T., Bork-Jensen, J., Brandslund, I., Christensen, C., Koistinen, H.A., Doney, A.S.F., Kinnunen, L., Esko, T., Farmer, A.J., Hakaste, L., Hodgkiss, D., Kravic, J., Lyssenko, V., Hollensted, M., Jorgensen, M.E., Jorgensen, T., Ladenvall, C., Justesen, J.M., Karajamaki, A., Kriebel, J., Rathmann, W., Lannfelt, L., Lauritzen, T., Narisu, N., Linneberg, A., Melander, O., Milani, L., Neville, M., Orho-Melander, M., Qi, L., Qi, Q.B., Roden, M., Rolandsson, O., Swift, A., Rosengren, A.H., Stirrups, K., Wood, A.R., Mihailov, E., Blancher, C., Carneiro, M.O., Maguire, J., Poplin, R., Shakir, K., Fennell, T., DePristo, M., de Angelis, M.H., Deloukas, P., Gjesing, A.P., Jun, G., Nilsson, P., Murphy, J., Onofrio, R., Thorand, B., Hansen, T., Meisinger, C., Hu, F.B., Isomaa, B., Karpe, F., Liang, L.M., Peters, A., Huth, C., O'Rahilly, S.P., Palmer, C.N.A., Pedersen, O., Rauramaa, R., Tuomilehto, J., Salomaa, V., Watanabe, R.M., Syvanen, A.C., Bergman, R.N., Bharadwaj, D., Bottinger, E.P., Cho, Y.S., Chandak, G.R., Chan, J.C.N., Chia, K.S., Daly, M.J., Ebrahim, S.B., Langenberg, C., Elliott, P., Jablonski, K.A., Lehman, D.M., Jia, W.P., Ma, R.C.W., Pollin, T.I., Sandhu, M., Tandon, N., Froguel, P., Barroso, I., Teo, Y.Y., Zeggini, E., Loos, R.J.F., Small, K.S., Ried, J.S., DeFronzo, R.A., Grallert, H., Glaser, B., Metspalu, A., Wareham, N.J., Walker, M., Banks, E., Gieger, C., Ingelsson, E., Im, H.K., Illig, T., Franks, P.W., Buck, G., Trakalo, J., Buck, D., Prokopenko, I., Magi, R., Lind, L., Farjoun, Y., Owen, K.R., Gloyn, A.L., Strauch, K., Tuomi, T., Kooner, J.S., Lee, J.Y., Park, T., Donnelly, P., Morris, A.D., Hattersley, A.T., Bowden, D.W., Collins, F.S., Atzmon, G., Chambers, J.C., Spector, T.D., Laakso, M., Strom, T.M., Bell, G.I., Blangero, J., Duggirala, R., Tai, E.S., McVean, G., Hanis, C.L., Wilson, J.G., Seielstad, M., Frayling, T.M., Meigs, J.B., Cox, N.J., Sladek, R., Lander, E.S., Gabriel, S., Burt, N.P., Mohlke, K.L., Meitinger, T., Groop, L., Abecasis, G., Florez, J.C., Scott, L.J., Morris, A.P., Kang, H.M., Boehnke, M., Altshuler, D. & McCarthy, M.I. 2016. The genetic architecture of type 2 diabetes. *Nature*, 536(7614):41-+.

Goh, K.-P. 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*. San Diego: Academic Press. p. 211-224).

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

Guey, L.T., Kravic, J., Melander, O., Burt, N.P., Laramie, J.M., Lyssenko, V., Jonsson, A., Lindholm, E., Tuomi, T., Isomaa, B., Nilsson, P., Almgren, P., Kathiresan, S., Groop, L., Seymour, A.B., Altshuler, D. & Voight, B.F. 2011. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet Epidemiol*, 35(4):236-246.

Hara, K., Fujita, H., Johnson, T.A., Yamauchi, T., Yasuda, K., Horikoshi, M., Peng, C., Hu, C., Ma, R.C., Imamura, M., Iwata, M., Tsunoda, T., Morizono, T., Shojima, N., So, W.Y., Leung, T.F., Kwan, P., Zhang, R., Wang, J., Yu, W., Maegawa, H., Hirose, H., consortium, D., Kaku, K., Ito, C., Watada, H., Tanaka, Y., Tobe, K., Kashiwagi, A., Kawamori, R., Jia, W., Chan, J.C., Teo, Y.Y., Shyong, T.E., Kamatani, N., Kubo, M., Maeda, S. & Kadowaki, T. 2014. Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet*, 23(1):239-246.

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

Hattingh, Z., Walsh, C.M., Bester, C.J. & Oguntibeju, O.O. 2008. Evaluation of energy and macronutrient intake of black women in Bloemfontein: A cross-sectional study. *African Journal of Biotechnology*, 7(22):4019-4024.

Hauner, H., Bechthold, A., Boeing, H., Bronstrup, A., Buyken, A., Leschik-Bonnet, E., Linseisen, J., Schulze, M., Strohm, D., Wolfram, G. & German Nutrition, S. 2012. Evidence-based guideline of the German Nutrition Society: carbohydrate intake and prevention of nutrition-related diseases. *Ann Nutr Metab*, 60 Suppl 1:1-58.

Hu, F.B. 2013. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. *Obesity Reviews*, 14(8):606-619.

Hu, F.B., Manson, J.E. & Willett, W.C. 2001. Types of dietary fat and risk of coronary heart disease: A critical review. *Journal of the American College of Nutrition*, 20(1):5-19.

Imamura, M., Maeda, S., Yamauchi, T., Hara, K., Yasuda, K., Morizono, T., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., Tsunoda, T., Kubo, M., Watada, H., Maegawa, H., Okada-Iwabu, M., Iwabu, M., Shojima, N., Ohshige, T., Omori, S., Iwata, M., Hirose, H., Kaku, K., Ito, C., Tanaka, Y., Tobe, K., Kashiwagi, A., Kawamori, R., Kasuga, M., Kamatani, N., Diabetes Genetics, R., Meta-analysis, C., Nakamura, Y. & Kadowaki, T. 2012. A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Hum Mol Genet*, 21(13):3042-3049.

Kahn, S.E., Cooper, M.E. & Del Prato, S. 2014. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet*, 383(9922):1068-1083.

Kahn, S.E., Prigeon, R.L., Mcculloch, D.K., Boyko, E.J., Bergman, R.N., Schwartz, M.W., Neifing, J.L., Ward, W.K., Beard, J.C., Palmer, J.P. & Porte, D. 1993. Quantification of the Relationship between Insulin Sensitivity and Beta-Cell Function in Human-Subjects - Evidence for a Hyperbolic Function. *Diabetes*, 42(11):1663-1672.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C.J., Fagerberg, B., Nielsen, J. & Backhed, F. 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*, 498(7452):99-103.

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. *Nat Genet*, 43(10):984-989.

Lander, E.S. & Botstein, D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121(1):185-199.

Lanktree, M.B., Hegele, R.A., Schork, N.J. & Spence, J.D. 2010. Extremes of unexplained variation as a phenotype: an efficient approach for genome-wide association studies of cardiovascular disease. *Circ Cardiovasc Genet*, 3(2):215-221.

Ley, S.H., Hamdy, O., Mohan, V. & Hu, F.B. 2014. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. *Lancet*, 383(9933):1999-2007.

Li, H., Gan, W., Lu, L., Dong, X., Han, X., Hu, C., Yang, Z., Sun, L., Bao, W., Li, P., He, M., Sun, L., Wang, Y., Zhu, J., Ning, Q., Tang, Y., Zhang, R., Wen, J., Wang, D., Zhu, X., Guo, K., Zuo, X., Guo, X., Yang, H., Zhou, X., Consortium, D., Consortium, A.-T.D., Zhang, X., Qi, L., Loos, R.J., Hu, F.B., Wu, T., Liu, Y., Liu, L., Yang, Z., Hu, R., Jia, W., Ji, L., Li, Y. & Lin, X. 2013. A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. *Diabetes*, 62(1):291-298.

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.

Li, Y.R. & Keating, B.J. 2014. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Med*, 6(10):91.

Ng, M.C. 2015. Genetics of Type 2 Diabetes in African Americans. *Curr Diab Rep*, 15(10):74.

MacIntyre, U.E., Kruger, H.S., Venter, C.S. & Vorster, H.H. 2002. Dietary intakes of an African population in different stages of transition in the North West Province, South Africa: the THUSA study. *Nutrition Research*, 22(3):239-256.

Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C., Prokopenko, I., Saleheen, D., Wang, X., Zeggini, E., Abecasis, G.R., Adair, L.S., Almgren, P., Atalay, M., Aung, T., Baldassarre, D., Balkau, B., Bao, Y., Barnett, A.H., Barroso, I., Basit, A., Been, L.F., Beilby, J., Bell, G.I., Benediktsson, R., Bergman, R.N., Boehm, B.O., Boerwinkle, E., Bonnycastle, L.L., Burt, N., Cai, Q., Campbell, H., Carey, J., Cauchi, S., Caulfield, M., Chan, J.C., Chang, L.C., Chang, T.J., Chang, Y.C., Charpentier, G., Chen, C.H., Chen, H., Chen, Y.T., Chia, K.S., Chidambaram, M., Chines, P.S., Cho, N.H., Cho, Y.M., Chuang, L.M., Collins, F.S., Cornelis, M.C., Couper, D.J., Crenshaw, A.T., van Dam, R.M., Danesh, J., Das, D., de Faire, U., Dedoussis, G., Deloukas, P., Dimas, A.S., Dina, C., Doney, A.S., Donnelly, P.J., Dorkhan, M., van Duijn, C., Dupuis, J., Edkins, S., Elliott, P., Emilsson, V., Erbel, R., Eriksson, J.G., Escobedo, J., Esko, T., Eury, E., Florez, J.C., Fontanillas, P., Forouhi,

N.G., Forsen, T., Fox, C., Fraser, R.M., Frayling, T.M., Froguel, P., Frossard, P., Gao, Y., Gertow, K., Gieger, C., Gigante, B., Grallert, H., Grant, G.B., Grrop, L.C., Groves, C.J., Grundberg, E., Guiducci, C., Hamsten, A., Han, B.G., Hara, K., Hassanali, N., Hattersley, A.T., Hayward, C., Hedman, A.K., Herder, C., Hofman, A., Holmen, O.L., Hovingh, K., Hreidarsson, A.B., Hu, C., Hu, F.B., Hui, J., Humphries, S.E., Hunt, S.E., Hunter, D.J., Hveem, K., Hydrie, Z.I., Ikegami, H., Illig, T., Ingelsson, E., Islam, M., Isomaa, B., Jackson, A.U., Jafar, T., James, A., Jia, W., Jockel, K.H., Jonsson, A., Jowett, J.B., Kadowaki, T., Kang, H.M., Kanoni, S., Kao, W.H., Kathiresan, S., Kato, N., Katulanda, P., Keinanen-Kiukaanniemi, K.M., Kelly, A.M., Khan, H., Khaw, K.T., Khor, C.C., Kim, H.L., Kim, S., Kim, Y.J., Kinnunen, L., Klopp, N., Kong, A., Korpi-Hyovalti, E., Kowlessur, S., Kraft, P., Kravic, J., Kristensen, M.M., Krithika, S., Kumar, A., Kumate, J., Kuusisto, J., Kwak, S.H., Laakso, M., Lagou, V., Lakka, T.A., Langenberg, C., Langford, C., Lawrence, R., Leander, K., Lee, J.M., Lee, N.R., Li, M., Li, X., Li, Y., Liang, J., Liju, S., Lim, W.Y., Lind, L., Lindgren, C.M., Lindholm, E., Liu, C.T., Liu, J.J., Lobbens, S., Long, J., Loos, R.J., Lu, W., Luan, J., Lyssenko, V., Ma, R.C., Maeda, S., Magi, R., Mannisto, S., Matthews, D.R., Meigs, J.B., Melander, O., Metspalu, A., Meyer, J., Mirza, G., Mihailov, E., Moebus, S., Mohan, V., Mohlke, K.L., Morris, A.D., Muhleisen, T.W., Muller-Nurasyid, M., Musk, B., Nakamura, J., Nakashima, E., Navarro, P., Ng, P.K., Nica, A.C., Nilsson, P.M., Njolstad, I., Nothen, M.M., Ohnaka, K., Ong, T.H., Owen, K.R., Palmer, C.N., Pankow, J.S., Park, K.S., Parkin, M., Pechlivanis, S., Pedersen, N.L., Peltonen, L., Perry, J.R., Peters, A., Pinidiyapathirage, J.M., Platou, C.G., Potter, S., Price, J.F., Qi, L., Radha, V., Rallidis, L., Rasheed, A., Rathman, W., Rauramaa, R., Raychaudhuri, S., Rayner, N.W., Rees, S.D., Rehnberg, E., Ripatti, S., Robertson, N., Roden, M., Rossin, E.J., Rudan, I., Rybin, D., Saaristo, T.E., Salomaa, V., Saltevo, J., Samuel, M., Sanghera, D.K., Saramies, J., Scott, J., Scott, L.J., Scott, R.A., Segre, A.V., Sehmi, J., Sennblad, B., Shah, N., Shah, S., Shera, A.S., Shu, X.O., Shuldiner, A.R., Sigurdsson, G., Sijbrands, E., Silveira, A., Sim, X., Sivapalaratnam, S., Small, K.S., So, W.Y., Stancakova, A., Stefansson, K., Steinbach, G., Steinthorsdottir, V., Stirrups, K., Strawbridge, R.J., Stringham, H.M., Sun, Q., Suo, C., Syvanen, A.C., Takayanagi, R., Takeuchi, F., Tay, W.T., Teslovich, T.M., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tikkanen, E., Trakalo, J., Tremoli, E., Trip, M.D., Tsai, F.J., Tuomi, T., Tuomilehto, J., Uitterlinden, A.G., Valladares-Salgado, A., Vedantam, S., Veglia, F., Voight, B.F., Wang, C., Wareham, N.J., Wennauer, R., Wickremasinghe, A.R., Wilsgaard, T., Wilson, J.F., Wiltshire, S., Winckler, W., Wong, T.Y., Wood, A.R., Wu, J.Y., Wu, Y., Yamamoto, K., Yamauchi, T., Yang, M., Yengo, L., Yokota, M., Young, R., Zabaneh, D., Zhang, F., Zhang, R., Zheng, W., Zimmet, P.Z., Altshuler, D., Bowden, D.W., Cho, Y.S., Cox, N.J., Cruz, M., Hanis, C.L., Kooner, J., Lee, J.Y., Seielstad, M., Teo, Y.Y., Boehnke, M., Parra, E.J., Chambers, J.C., Tai, E.S., McCarthy, M.I., Morris, A.P., Replication, D.G., Meta-analysis, C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., South Asian Type 2 Diabetes, C., Mexican American Type 2 Diabetes, C. & Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples, C. 2014. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*, 46(3):234-244.

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*, 14:644-2164-2114-2644.

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.

McEvoy, C.T., Cardwell, C.R., Woodside, J.V., Young, I.S., Hunter, S.J. & McKinley, M.C. 2014. A posteriori dietary patterns are related to risk of type 2 diabetes: findings from a systematic review and meta-analysis. *J Acad Nutr Diet*, 114(11):1759-1775 e1754.

Medici, F., Hawa, M., Ianari, 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.

Mitri, J., Muraru, M.D. & Pittas, A.G. 2011. Vitamin D and type 2 diabetes: a systematic review. *European Journal of Clinical Nutrition*, 65(9):1005-1015.

Morris, A.P. 2011. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol*, 35(8):809-822.

Moskal, A., Pisa, P.T., Ferrari, P., Byrnes, G., Freisling, H., Boutron-Ruault, M.C., Cadeau, C., Nailler, L., Wendt, A., Kuhn, T., Boeing, H., Buijsse, B., Tjonneland, A., Halkjaer, J., Dahm, C.C., Chiuve, S.E., Quiros, J.R., Buckland, G., Molina-Montes, E., Amiano, P., Castano, J.M.H., Gurrea, A.B., Khaw, K.T., Lentjes, M.A., Key, T.J., Romaguera, D., Vergnaud, A.C., Trichopoulou, A., Bamia, C., Orfanos, P., Palli, D., Pala, V., Tumino, R., Sacerdote, C., de Magistris, M.S., Bueno-de-Mesquita, H.B., Ocke, M.C., Beulens, J.W.J., Ericson, U., Drake, I., Nilsson, L.M., Winkvist, A., Weiderpass, E., Hjartaker, A., Riboli, E. & Slimani, N. 2014. Nutrient Patterns and Their Food Sources in an International Study Setting: Report from the EPIC Study. *PloS one*, 9(6).

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.

Ng, M.C.Y. 2015. Genetics of Type 2 Diabetes in African Americans. *Current Diabetes Reports*, 15(10).

Nolan, C.J., Damm, P. & Prentki, M. 2011. Type 2 diabetes across generations: from pathophysiology to prevention and management. *Lancet*, 378(9786):169-181.

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., Consortium, D., Investigators, M., 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., Burt, 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., Lieve, A., Lindgren, C.M., Lyssenko, V., Marre, M., Meitinger, T., Midtjell, 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., Gyllenstein, 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., Roccascaccia, 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., Ruukonen, 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.

Perry, J.R., Voight, B.F., Yengo, L., Amin, N., Dupuis, J., Ganster, M., Grallert, H., Navarro, P., Li, M., Qi, L., Steinthorsdottir, V., Scott, R.A., Almgren, P., Arking, D.E., Aulchenko, Y., Balkau, B., Benediktsson, R., Bergman, R.N., Boerwinkle, E., Bonnycastle, L., Burt, N.P., Campbell, H., Charpentier, G., Collins, F.S., Gieger, C., Green, T., Hadjadj, S., Hattersley, A.T., Herder, C., Hofman, A., Johnson, A.D., Kottgen, A., Kraft, P., Labrune, Y., Langenberg, C., Manning, A.K., Mohlke, K.L., Morris, A.P., Oostra, B., Pankow, J., Petersen, A.K., Pramstaller, P.P., Prokopenko, I., Rathmann, W., Rayner, W., Roden, M., Rudan, I., Rybin, D., Scott, L.J., Sigurdsson, G., Sladek, R., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Uitterlinden, A.G., Vivequin, S., Weedon, M.N., Wright, A.F., MAGIC Consortium, D., Consortium, G., Hu, F.B., Illig, T., Kao, L., Meigs, J.B., Wilson, J.F., Stefansson, K., van Duijn, C., Altschuler, D., Morris, A.D., Boehnke, M., McCarthy, M.I., Froguel, P., Palmer, C.N., Wareham, N.J., Groop, L., Frayling, T.M. & Cauchi, S. 2012. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet*, 8(5):e1002741.

Pirie, F.J., Motala, A.A., Pegoraro, R.J., Paruk, I.M., 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. *Afr J Diab Med*, 18(1):12-16.

Popkin, B.M., Adair, L.S. & Ng, S.W. 2012. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev*, 70(1):3-21.

Prasad, R.B. & Groop, L. 2015. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes (Basel)*, 6(1):87-123.

Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Pare, G., Sun, Q., Girman, C.J., Laurie, C.C., Mirel, D.B., Manolio, T.A., Chasman, D.I., Boerwinkle, E., Ridker, P.M., Hunter, D.J., Meigs, J.B., Lee, C.H., Hu, F.B., van Dam, R.M., Meta-Analysis of, G., Insulin-related traits, C., Diabetes Genetics, R. & Meta-analysis, C. 2010. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet*, 19(13):2706-2715.

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

Ramsay, M., Sankoh, O., as members of the, A.W.I.G.s. & the, H.A.C. 2016. African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent. *Int J Epidemiol*, 45(2):305-308.

Reaven, G.M. 1988. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes*, 37(12):1595-1607.

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

Risch, N. 1990. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet*, 46(2):242-253.

Riserus, U., Willett, W.C. & Hu, F.B. 2009. Dietary fats and prevention of type 2 diabetes. *Prog Lipid Res*, 48(1):44-51.

Salmeron, J., Hu, F.B., Manson, J.E., Stampfer, M.J., Colditz, G.A., Rimm, E.B. & Willett, W.C. 2001. Dietary fat intake and risk of type 2 diabetes in women. *Am J Clin Nutr*, 73(6):1019-1026.

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 glycotxin intake. *Diabetes*, 54(8):2314-2319.

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.

Saxena, R., Voight, B.F., Lyssenko, V., Burt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J., Hughes, T.E., Groop, L., Altshuler, D., Almgren, P., Florez, J.C., Meyer, J., Ardlie, K., Bengtsson Bostrom, K., Isomaa, B., Lettre, G., Lindblad, U., Lyon, H.N., Melander, O., Newton-Cheh, C., Nilsson, P., Orho-Melander, M., Rastam, L., Speliotes, E.K., Taskinen, M.R., Tuomi, T., Guiducci, C., Berglund, A., Carlson, J., Gianniny, L., Hackett, R., Hall, L., Holmkvist, J., Laurila, E., Sjogren, M., Sterner, M., Surti, A., Svensson, M., Svensson, M., Tewhey, R., Blumenstiel, B., Parkin, M., Defelice, M., Barry, R., Brodeur, W., Camarata, J., Chia, N., Fava, M., Gibbons, J., Handsaker, B., Healy, C., Nguyen, K., Gates, C., Sougnez, C., Gage, D., Nizzari, M., Gabriel, S.B., Chirn, G.W., Ma, Q., Parikh, H., Richardson, D., Ricke, D., Purcell, S., Diabetes Genetics Initiative of Broad Institute of, H., Mit, L.U. & Novartis Institutes of BioMedical, R. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316(5829):1331-1336.

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.J., Swift, A.J., Narisu, H., Hu, T., Pruim, R., Xiao, R., Li, X.Y., 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.

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.

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.

Smith, J.A., Ware, E.B., Middha, P., Beacher, L. & Kardia, S.L. 2015. Current Applications of Genetic Risk Scores to Cardiovascular Outcomes and Subclinical Phenotypes. *Curr Epidemiol Rep*, 2(3):180-190.

Steyn, N.P. & Temple, N.J. 2012. Evidence to support a food-based dietary guideline on sugar consumption in South Africa. *BMC Public Health*, 12.

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

Tinker, L.F., Bonds, D.E., Margolis, K.L., Manson, J.E., Howard, B.V., Larson, J., Perri, M.G., Beresford, S.A., Robinson, J.G., Rodriguez, B., Safford, M.M., Wenger, N.K., Stevens, V.J., Parker, L.M. & Women's Health, I. 2008. Low-fat dietary pattern and risk of treated diabetes mellitus in postmenopausal women: the Women's Health Initiative randomized controlled dietary modification trial. *Arch Intern Med*, 168(14):1500-1511.

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.

Vassy, J.L., Hivert, M.F., Porneala, B., Dauriz, M., Florez, J.C., Dupuis, J., Siscovick, D.S., Fornage, M., Rasmussen-Torvik, L.J., Bouchard, C. & Meigs, J.B. 2014. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes*, 63(6):2172-2182.

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., Burt, 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., Lieveise, 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.K., 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., Gyllenstein, 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.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. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*, 42(7):579-589.

Vorster, H.H., Kruger, A. & Margetts, B.M. 2011. The Nutrition Transition in Africa: Can It Be Steered into a More Positive Direction? *Nutrients*, 3(4):429-441.

Vrieze, A., Van Nood, E., Holleman, F., Salojarvi, J., Kootte, R.S., Bartelsman, J.F., Dallinga-Thie, G.M., Ackermans, M.T., Serlie, M.J., Oozeer, R., Derrien, M., Druesne, A., Van Hylckama Vlieg, J.E., Bloks, V.W., Groen, A.K., Heilig, H.G., Zoetendal, E.G., Stroes, E.S., de Vos, W.M., Hoekstra, J.B. & Nieuwdorp, M. 2012. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology*, 143(4):913-916 e917.

Wang, X., Strizich, G., Hu, Y., Wang, T., Kaplan, R.C. & Qi, Q. 2015. Genetic markers of type 2 diabetes: Progress in genome-wide association studies and clinical application for risk prediction. *J Diabetes*.

Weyer, C., Bogardus, C., Mott, D.M. & Pratley, R.E. 1999. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest*, 104(6):787-794.

WHO. 2011. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation. Geneva.

Willett, W. 1998. Nutritional epidemiology. New York: Oxford University Press.

Wu, J.H., Micha, R., Imamura, F., Pan, A., Biggs, M.L., Ajaz, O., Djousse, L., Hu, F.B. & Mozaffarian, D. 2012. Omega-3 fatty acids and incident type 2 diabetes: a systematic review and meta-analysis. *Br J Nutr*, 107 Suppl 2:S214-227.

Yako, Y.Y., Guewo-Fokeng, M., Balti, E.V., Bouatia-Naji, N., Matsha, T.E., Sobngwi, E., Erasmus, R.T., Echouffo-Tcheugui, J.B. & Kengne, A.P. 2016. Genetic risk of type 2 diabetes in populations of the African continent: A systematic review and meta-analyses. *Diabetes Res Clin Pract*, 114:136-150.

Yalow, R.S. & Berson, S.A. 1960. Immunoassay of endogenous plasma insulin in man. *J Clin Invest*, 39:1157-1175.

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. *Nat Genet*, 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. *Nat Genet*, 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. *Nat Genet*, 40(5):638-645.

Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M., Barrett, J.C., Shields, B., Morris, A.P., Ellard, S., Groves, C.J., Harries, L.W., Marchini, J.L., Owen, K.R., Knight, B., Cardon, L.R., Walker, M., Hitman, G.A., Morris, A.D., Doney, A.S., Wellcome Trust Case Control, C., McCarthy, M.I. & Hattersley, A.T. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316(5829):1336-1341.

Zhao, Z.X., Li, S.Y., Liu, G.J., Yan, F.F., Ma, X.L., Huang, Z.Y. & Tian, H.M. 2012. Body Iron Stores and Heme-Iron Intake in Relation to Risk of Type 2 Diabetes: A Systematic Review and Meta-Analysis. *PloS one*, 7(7).

CHAPTER 3

The predictive utility of a genetic risk score of common variants associated with T2D in a black South African population

Predictive utility of a genetic risk score of common variants associated with type 2 diabetes in a black South African population

Tinashe Chikowore^{1*}, Tertia van Zyl¹, Edith J.M. Feskens² and Karin R. Conradie¹

¹Centre for Excellence in Nutrition, North-West University, Potchefstroom, North West Province, 2520, South Africa; ² Wageningen University, Division of Human Nutrition, P.O. Box 17, 6700 AA Wageningen, The Netherlands

Correspondence:

Tinashe Chikowore

Centre of Excellence for Nutrition

North West University

Potchefstroom, North West Province, 2520

South Africa

Email: tinashedoc@gmail.com;

Phone: 0027781169789

Fax: 0027182992464

Abstract

Aims: To determine the predictive utility of polygenic risk scores of common variants associated with type 2 diabetes derived from the European and Asian ethnicities among a black South African population.

Method: Our study was a case-control study nested within the Prospective Urban and Rural Epidemiological (PURE) study of 178 male and female cases, matched for age and gender with 178 controls. Four types of genetic risk scores (GRS) were developed from 66 selected SNPs. These comprised of beta cell related variants (GRS_b), variants which had significant associations with T2D in our study (GRS_n), variants from the trans-ethnic meta-analysis (GRS_{trans}) and all the 66 selected SNPs (GRS_t).

Results: Of the GRS's, only GRS_n was associated with increased risk of T2D as indicated by an OR (95CI) of 1.21 (1.02-1.43) p-value = 0.015. Stratified analysis of age and BMI, indicated the GRS_n to be significantly associated with T2D among the non-obese and participants less than 50 years. The area under the ROC of the T2D risk factors only was 0.652 (p value < 0.001) and with the addition of GRS_n it was 0.665 (p value < 0.001).

Conclusions: The GRS of European and Asian derived variants have limited clinical utility in the black South African population. The inclusion of population specific variants in the GRS is pivotal.

Keywords: GRS, polygenic risk score, clinical use, BMI

1. Introduction

The prevalence of the type 2 diabetes (T2D) disease burden is increasing globally and South Africans have not been spared [1, 2]. Approximately 2 million South Africans live with T2D, with about 115 000 new cases occurring annually [1]. However, limited evidence exists about the etiology of T2D, which is hampering the development of effective therapeutic and preventative interventions to curb this growing burden [3]. Added to this, there is an urgent need to scale up diabetes screening, as it is estimated that 30-80% of the new T2D cases remain undiagnosed and that even in those that are diagnosed, 20% of them would have developed irreversible complications by the time of diagnoses. [4].

The strong genetic component and heritability of T2D which is indicated by concordance rates of 79% among monozygotic twins provides an opportunity to identify high risk individuals well before they have developed the disease [5, 6]. The genetic risk score (GRS) is an approach that combines multiple SNP information into a single metric representing the genetic predisposition for an individual to a particular phenotype or disease trait [7]. The GRS has been depicted to be more predictive in participants who are non-obese and not older than 50 years of age [8]. Using the GRS to identify high risk individuals while still young and without risk factors such as obesity being present, is therefore very relevant. Although individual variants have been depicted to have small effect sizes, there is a growing interest that a combination of these variants in the form of a GRS might be more useful in predicting disease susceptibility [9]. As opposed to the genome wide association studies (GWAS) which require large sample sizes to identify

significant variants, the GRS were noted not to vary as a result of sample size and study design [10]. Thereby, indicating the GRS approach relevant for evaluating genetic risk in modest samples sizes that are common among study settings in Africa due to resource constraints.

Numerous studies have been done among Caucasians to evaluate the effectiveness of the GRS alone and in combination with conventional T2D risk factors such as age, sex and body mass index (BMI) [7, 9, 10]. However, it has been noted to only modestly improve the predictive ability of the conventional risk factors of T2D among the Caucasians. Little is known of the association of GRS of common variants derived from European and Asian ethnicities with T2D in Africa [7, 10]. However, the GRS of European derived common variants was reported to be significantly associated with T2D susceptibility among Caucasians and African Americans regardless that individual common variants associated with T2D in these ethnic groups are different [9]. Linkage disequilibrium (LD) differences among different ethnic groups have been identified as the main factor for not being able to replicate the common variants associations [11]. However, it seems that the GRS is not limited by LD and this aspect needs to be further investigated in other populations of African ancestry which have not been able to replicate the association of common variants with T2D [9]. This study was done to determine the GRS constructed from common variants derived from European and Asian ethnicities associated with T2D, its predictive utility alone and in combination with other conventional risk factors in a black South African population of Sestwana descent.

2.0 Material and Methods

2.1 Study Participants

One hundred and seventy eight cases and 178 controls were recruited for this study from the South African arm of the PURE study as described elsewhere [12]. The 178 cases were matched for gender and age with 178 controls. The cases were identified as having blood glucose levels above 7.8 mmol.L^{-1} while the controls had normal glucose levels (less than 7.8 mmol.L^{-1}) after a two-hour oral glucose tolerance test (OGTT) were administered to each participant. The participants completed written informed consents to participate in the study. The study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the North-West University, Potchefstroom Campus, with ethics number NWU-00016-10-A1.

2.2 Biochemical measurements

The research participants who were in a fasting state (at least 10 hours with no food or beverages, including water) had their blood glucose levels measured by the PURE research team. The fasting glucose levels were measured using the SYNCHRON® System from fluoride plasma. Thereafter, the OGTT was performed by giving the participants a 75 gram glucose load dissolved in 300 mL of water to ingest and post-prandial glucose measurement from were determined from withdrawn blood at 30, 60, 90 and 120 minutes as described in detail elsewhere [12]. The Bio-Rad D-10™, HbA1c kit (Bio-Rad Laboratories, Inc., Hercules, France), which operates via cation exchange high

performance liquid chromatography was used to assess HbA1c levels from whole blood EDTA samples.

2.3 BMI and blood pressure

The height was measured using a stadiometer (IP 1465, Invicta, and London, UK) and weight was assessed using portable electronic scales (Precision Health Scale, A&D Tokyo, Japan). The measurements were done in duplicate according to standardised protocols and the mean was recorded. BMI was computed as, $BMI = \text{weight (kilograms)} / \text{height (metres)}^2$. The participants were allowed a 10-minute rest period, after which BP was measured twice, after a duration of 5 minutes. Blood pressure was measured on the upper right arm while the participants were seated upright and the arm supported at heart level. The OMRON HEM-757 (Omron Healthcare, Kyoto, Japan) was used to measure systolic and diastolic blood pressure. Appropriately sized cuffs were used for the obese participants.

2.4 DNA preparation and Genotyping

The Qiagen Flexigene DNA extraction kit (catalogue number 51206, QIAGEN Pty. Ltd., Australia) was used to extract genomic DNA from citrated treated buffy coat. Seventy-nine SNPs associated with T2D were selected from literature and genotyped using the GoldenGate Assay with VeraCode technology (Illumina Inc., San Diego, CA, USA) on the BeadXpress platform (Illumina Inc., San Diego, CA, USA). After quality control 66 SNPs remained for association tests. The details regarding the DNA preparation, genotyping and quality control are described in more detail elsewhere [12].

2.5 Construction of the GRS

A simple count (unweighted) method was used to calculate the GRS which assumes the additive genetic model for each SNP. However, the weighted GRS approach which takes into account the different effects sizes of SNPs on T2D risk was not used due to the absence of GWAS studies for T2D in African populations, from which it is recommended to select effect sizes for constructing the weighted GRS [13]. Effect size of the risk alleles could not be selected from GWAS studies among the Asian and European populations, because of the extreme differential effects of the T2D risk alleles among different population groups that have been noted, for instance, only 10% of the 103 known T2D variants having the same risk allele and direction of effect size in African populations [14-16]. The additive genetic model performs well when the true genetic model is unknown or wrongly indicated [15]. The simple count method also assumes that the SNPs under investigation have similar contribution to T2D risk [15]. The risk alleles of the selected 66 T2D SNPs used in this study were identical to those from which the SNPs were initially discovered as in Table 1. The genotypes of the selected SNPs were assigned a weight of 0, 1 and 2 depending on the number of risk alleles. The 66 SNPs had a proportion of 1, 5 % of missing values. Due to the limited number of SNPs in our study, which were not genotyped using a SNP array and not strongly linked, the LD methods for imputation were not used. The 66 SNPs were deemed not dense enough compared to SNP arrays that comprise of 0.3 to 5million SNPs, to form a genotyping scaffold that would enable accurate imputation using LD methods[17]. Imputation accuracy using LD methods has been reported to be adversely affected by SNP density [18]. The series mean imputation method was selected as it had been reported in similar studies with limited SNPs [19, 20].

Missing genotype data was imputed with the mean score of the respective SNP. The GRS was determined by summing the scores of the selected SNPs.

Four categories of GRS's from 66 SNPs were calculated to determine the score which was significantly associated with increased risk of T2D in the black South African population as indicated in Table 1. The first GRS was made up of the variants which were previously reported to have statistical significant associations in the black South African population and it was denoted as GRS_n [12]. Rs8050136, rs831571, rs7542900 and rs1436955 were the variants used to calculate GRS_n. The SNPs reported to be related with beta cell dysfunction among the selected 66 SNPs were selected and a GRS was computed termed GRS_{beta}. Rs1436955, rs1387153, rs13266634, rs7754840 and rs7903146 were used to construct GRS_{beta} [21, 22]. The mechanistic effect of these five SNPs in relation to beta cell dysfunction has been characterised [22]. However, most of the other SNPs have been noted to be in intergenic regions which makes it difficult to characterise the mechanisms underlying their associations with T2D risk [22]. Among the 66 SNPs investigated in this study were variants which had significant associations in GWAS trans-ethnic meta-analysis for T2D, these were used to compute the GRS termed GRStrans [23]. The GRStrans consisted of rs1387153, rs13266634, rs7754840, rs7903146, rs1111875, rs4402960, rs564398, rs8042680, rs864745 variants. The final GRS comprised of all the 66 SNPs and it was depicted as GRSt.

Table 1 Genetic risk scores and their related variants

GRS	Nearest Loci	SNP ID	Risk Allele	Other allele	Study used to designate the SNPs into a selected category
GRSn	<i>FTO</i>	rs8050136	A	C	Chikowore et al, 2015 [12]
	<i>PSMD6</i>	rs831571	C	T	
	<i>SLC44A3, F3</i>	rs7542900	C	T	
	<i>C2CD4B</i>	rs1436955	C	T	
GRSbeta	<i>MTNR1B</i>	rs1387153	T	C	Qi and Hu, 2012 [21]
	<i>TCF7L2</i>	rs7903146	T	C	
	<i>C2CD4B</i>	rs1436955	C	T	
	<i>SLC30A8</i>	rs1326634	T	C	
	<i>CDKAL1</i>	rs7754840	C	A	
GRStrans	<i>HHEX</i>	rs1111875	C	T	Morris et al, 2012 [23]
	<i>IGF2BP2</i>	rs4402960	T	G	
	<i>CDKN2A, CD-KN2B</i>	rs564398	T	C	
	<i>PRC1</i>	rs8042680	A	C	
	<i>JAZF1</i>	rs864745	T	C	
	<i>MTNR1B</i>	rs1387153	T	C	
	<i>TCF7L2</i>	rs7903146	T	C	
	<i>SLC30A8</i>	rs1326634	T	C	
	<i>CDKAL1</i>	rs7754840	C	A	
GRSt	<i>LPIN2</i>	rs10460009	C	T	Sim et al., 2011 [24]
	<i>C6orf57</i>	rs1048886	G	A	Sim et al., 2011 [24]
	<i>TCERG1L</i>	rs10741243	G	C	Sim et al., 2011 [24]
	<i>NOTCH2, ADAM30</i>	rs10923931	T	G	Zeggini et al., 2008 [25]
	<i>HHEX</i>	rs1111875	C	T	Scott et al., 2007 [26]
	<i>DCD</i>	rs1153188	A	T	Zeggini et al., 2008 [25]
	<i>ZFAND6</i>	rs11634397	A	G	Voight et al., 2010 [27]
	<i>LOC72901, CETN3</i>	rs12518099	G	A	Rung et al., 2009 [28]
	<i>CDC123, CAMK1D</i>	rs12779790	G	A	Zeggini et al., 2008 [25]
	<i>SLC30A8</i>	rs1326634	T	C	Scott et al., 2007 [26]
	<i>CHCHD9</i>	rs13292136	C	T	Voight et al., 2010 [27]
	<i>SPRY2</i>	rs1359790	G	A	Shu et al., 2010 [29]
	<i>MTNR1B</i>	rs1387153	T	C	Voight et al., 2010 [27]
	<i>C2CD4B</i>	rs1436955	C	T	Shu et al., 2010 [29]

<i>HMGA2</i>	rs1531343	C	G	Voight et al., 2010 [27]
<i>ST6GAL1</i>	rs16861329	C	T	Kooner et al., 2011 [30]
<i>CMIP</i>	rs16955379	C	T	Cho et al., 2011 [31]
<i>SYN2, PPARG</i>	rs17036101	G	A	Zeggini et al., 2008 [25]
<i>PTPRD</i>	rs17584499	T	C	Tsai et al., 2010 [32]
<i>WWOX</i>	rs17797882	T	C	Cho et al., 2011 [31]
<i>AP3S2</i>	rs2028299	C	A	Kooner et al., 2011 [30]
<i>KCNQ1</i>	rs231362	G	A	Voight et al., 2010 [27]
<i>BCL11A</i>	rs243021	A	G	Voight et al., 2010 [27]
<i>GALNTL4,</i> <i>LOC729013</i>	rs2722769	C	G	Palmer et al., 2012 [33]
<i>HUNK</i>	rs2833610	A	G	Sim et al., 2011 [24]
<i>LOC64673,</i> <i>IRS1</i>	rs2943641	C	T	Rung et al., 2009 [28]
<i>NR</i>	rs358806	A	C	WTCC Consortium, 2007
<i>PLS1</i>	rs3773506	C	G	Sim et al., 2011 [24]
<i>SRR</i>	rs391300	C	T	Tsai et al., 2010 [32]
<i>GRB14</i>	rs3923113	A	C	Kooner et al., 2011 [30]
<i>IGF2BP2</i>	rs4402960	T	G	Scott et al., 2007 [26]
<i>HNF1B</i>	rs4430796	G	A	Voight et al., 2010 [27]
<i>ZBED3</i>	rs4457053	G	A	Voight et al., 2010 [27]
<i>ADAMTS9</i>	rs4607103	C	T	Zeggini et al., 2008 [25]
<i>WFS1,</i> <i>PPP2R2C</i>	rs4689388	A	G	Rung et al. 2009 [28]
<i>PAPL</i>	rs472265	G	A	Sim et al., 2011 [24]
<i>HNF4A</i>	rs4812829	A	G	Kooner et al., 2011 [30]
<i>CDKN2A, CD-</i> <i>KN2B</i>	rs564398	T	C	Zeggini et al., 2007 [34]
<i>DUSP9</i>	rs5945326	G	A	Voight et al., 2010 [27]
<i>GCC1, PAX4</i>	rs6467136	G	A	Cho et al., 2011 [31]
<i>KIF11</i>	rs6583826	G	A	Sim et al., 2011 [24]
<i>MAEA</i>	rs6815464	C	G	Cho et al., 2011 [31]
<i>CDKN2B</i>	rs7018475	G	T	Huang et al., 2012 [35]
<i>GLIS3</i>	rs7041847	A	G	Cho et al., 2011 [31]
<i>TMEM45B,</i> <i>BARX2</i>	rs7107217	C	A	Palmer et al., 2012 [33]
<i>C2CD4A,</i> <i>C2CD4B</i>	rs7172432	A	G	Yamauchi et al., 2010 [36]
<i>HMG20A</i>	rs7178572	G	A	Kooner et al., 2011 [30]
<i>C14orf70</i>	rs730570	G	A	Parra et al., 2011 [37]
<i>SLC44A3, F3</i>	rs7542900	C	T	Palmer et al., 2012 [33]
<i>RBM43,</i> <i>RND3</i>	rs7560163	C	G	Palmer et al., 2012 [33]
<i>IRS1</i>	rs7578326	A	G	Voight et al., 2010 [27]

<i>THADA</i>	rs7578597	T	C	Zeggini et al., 2008 [25]
<i>PEX5L</i>	rs7630877	A	G	Sim et al., 2011 [24]
<i>ACHE</i>	rs7636	A	G	Sim et al., 2011 [24]
<i>CDKAL1</i>	rs7754840	C	A	Scott et al., 2007 [26]
<i>TCF7L2</i>	rs7903146	T	C	Saxena et al 2007 [38]
<i>HNF1A</i>	rs7957197	T	A	Voight et al., 2010 [27]
<i>TSPAN8,</i> <i>LGR5</i>	rs7961581	C	T	Zeggini et al., 2008 [25]
<i>PRC1</i>	rs8042680	A	C	Voight et al., 2010 [27]
<i>FTO</i>	rs8050136	A	C	Zeggini et al., 2007 [34]
<i>PSMD6</i>	rs831571	C	T	Cho et al., 2011 [31]
<i>JAZF1</i>	rs864745	T	C	Zeggini et al., 2008 [25]
<i>TP53INP1</i>	rs896854	T	C	Voight et al., 2010 [27]
<i>ZFAND3</i>	rs9470794	C	T	Cho et al., 2011 [31]
<i>VEGFA</i>	rs9472138	T	C	Zeggini et al., 2008 [25]
<i>KLF14</i>	rs972283	G	A	Voight et al., 2010 [27]

2.6 Statistical Analysis

The statistical tests were done using SPSS version 21 and Plink version 2.050 [39]. Q-Q plots were used to evaluate the normality of the continuous variables. Chi square tests were used to assess the differences of the categorical variables among the cases and controls. The adherence of the selected variants to the Hardy Weinberg assumptions were evaluated as described previously [12]. The student t tests were used to compare the continuous variables among the cases and controls. Logistic regression was used to evaluate the association of the GRS with type 2 diabetes risk while adjusting for age, sex, BMI and urbanization. The GRS which was the significantly associated with increased T2D risk was selected. The predictive utility of the GRS was assessed by reviewing its improvement on discriminating between the cases and controls using the receiver operated characteristic curves (ROC) and the corresponding areas under the curves (AUC). The ROC of the logistic regression probabilities including conventional T2D risk

factors age, sex, BMI and systolic blood pressure were computed alone and in combination with the GRS associated with increased T2D risk. A p-value of 0.05 was used to denote statistical significance.

3.0 Results

The cases evaluated in the study were significantly more overweight than the controls with a BMI of 26.1kg/m² compared to the controls' BMI of 22.92 kg/m² (p-value<0.001). A greater proportion of the cases (68%, p value = 0.001) resided in urban areas as opposed to the controls as illustrated in Table 2. Only GRSn was significantly associated with increased risk of T2D as indicated by an odds ratio (OR) of 1.21 (p-value = 0.032) as illustrated in Table 3.

Table 2 Descriptive statistics of the participants

Variable	Cases	Controls	P value
Gender n (Male/Females)	39/139	39/139	
Mean age(years)	42.88 ± 4.90	42.51 ± 5.16	0.481
BMI (kg/m ²)	26.10 ± 7.82	22.92 ± 6.02	<0.001
Systolic blood pressure (mmHg)	124.14 ± 18.72	122.02 ± 19.88	0.303
Diastolic blood pressure(mmHg)	84.87 ± 14.16	82.61 ± 13.92	0.132
Fasting glucose (mmol.L ⁻¹)	4.92 ± 0.78	4.72 ± 0.66	0.011
HbA1C % (mmol/mol)	5.6 (38) ± 0.5	5.5 (37) ± 0.4	0.022
Urban resident's n (%)	68 (38.2)	39 (21.9)	0.001
2 hr Glucose (mmol.L ⁻¹)	9.30 ± 1.40	6.59 ± 0.78	<0.001

HbA1C measure indicated in NGSP% and IFCC mmol/mol; n= number.

Table 3 Associations of the selected GRS's with risk for type 2 diabetes

Variable	GRSn			GRSbeta			GRStrans			GRSt		
	OR	95 (CI)	P value	OR	95 (CI)	P value	OR	95 (CI)	P value	OR	95 (CI)	P value
Age (years)	1.01	0.97 - 1.06	0.597	1.01	0.97 – 1.06	0.597	1.01	0.97 -1.06	0.627	1.01	0.97 -1.05	0.671
Gender*	1.04	0.61 – 1.76	0.890	1.07	0.63 – 1.80	0.856	1.12	0.66 -1.90	0.665	1.10	0.65 – 1.86	0.726
BMI (kg/m ²)	1.07	1.04 - 1.10	<0.001	1.07	1.03 – 1.10	<0.001	1.07	1.04 – 1.10	<0.001	1.07	1.03 – 1.10	<0.001
GRS*	1.21	1.02 – 1.43	0.032	0.91	0.78 – 1.07	0.266	0.84	0.73 – 0.92	0.020	0.94	0.90 -0.99	0.019

Gender=male is reference category; GRS* = category specific genetic risk score respective of the column headings; CI = confidence interval; kg/m² =kilograms per meter squared; OR=odds ratio.

However, the GRStrans and GRSt were significantly less predictive for T2D risk as shown by OR's of 0.84 (p-value = 0.02) and 0.94 (p-value = 0.019) respectively as illustrated in Table 3. The interactions of the GRSn with Age and BMI were separately evaluated and the GRSn was noted to have significant interactions of P = 0.033 and P = <0.001 with age and BMI respectively. Further explorations indicated the GRSn to be more predictive of risk in participant less than 50 years (OR 1.23; p value = 0.023) and non-obese (OR 1.21; p value = 0.047) compared to those 50 years or more (OR 0.89; p value = 0.737) and the obese (OR 1.18; p value 0.465), respectively as illustrated in Table 4.

Table 4 Age and BMI stratified associations of the GRSn with risk for type 2 diabetes

Variable	Obese		Non-Obese		Less than 50 years		50 years or more	
	OR (95 CI)	P-value	OR (95 CI)	P value	OR (95 CI)	P value	OR (95 CI)	P value
Age (years)	1.12 (1.00 - 1.26)	0.056	1.00 (0.95 -1.04)	0.844	1.00 (0.95 - 1.05)	0.991	1.05 (0.75 - 1.48)	0.781
Gender*	1.18 (0.36 - 3.88)	0.791	0.91 (0.50 - 1.66)	0.762	1.02 (0.59 - 1.77)	0.947	0.59 (0.09 - 3.84)	0.58
BMI (kg/m²)	1.01 (0.90 -1.13)	0.866	1.07 (1.00 - 1.14)	0.053	1.07 (1.03 - 1.10)	<0.001	1.12 (0.96 - 1.32)	0.159
GRSn	1.18 (0.76 -1.82)	0.465	1.21 (1.00 - 1.46)	0.047	1.23 (1.03-1.47)	0.023	0.89 (0.46 - 1.73)	0.737

Gender=male is reference category; GRSn = genetic risk score for the significant SNPs associated with type 2 diabetes in this study; CI = confidence interval; kg/m² = kilograms per meter squared; OR=odds ratio.

The AUC of the ROC was 0.652 ± 0.029 p value<0.001 for the conventional risk factors and it was modestly increased to 0.665 ± 0.029 p value<0.001 as a result of the addition of the GRSn as illustrated in Figure 1. Thereby, indicating the modest improvement in the predictiveness of conventional risk factors with the addition of the GRSn

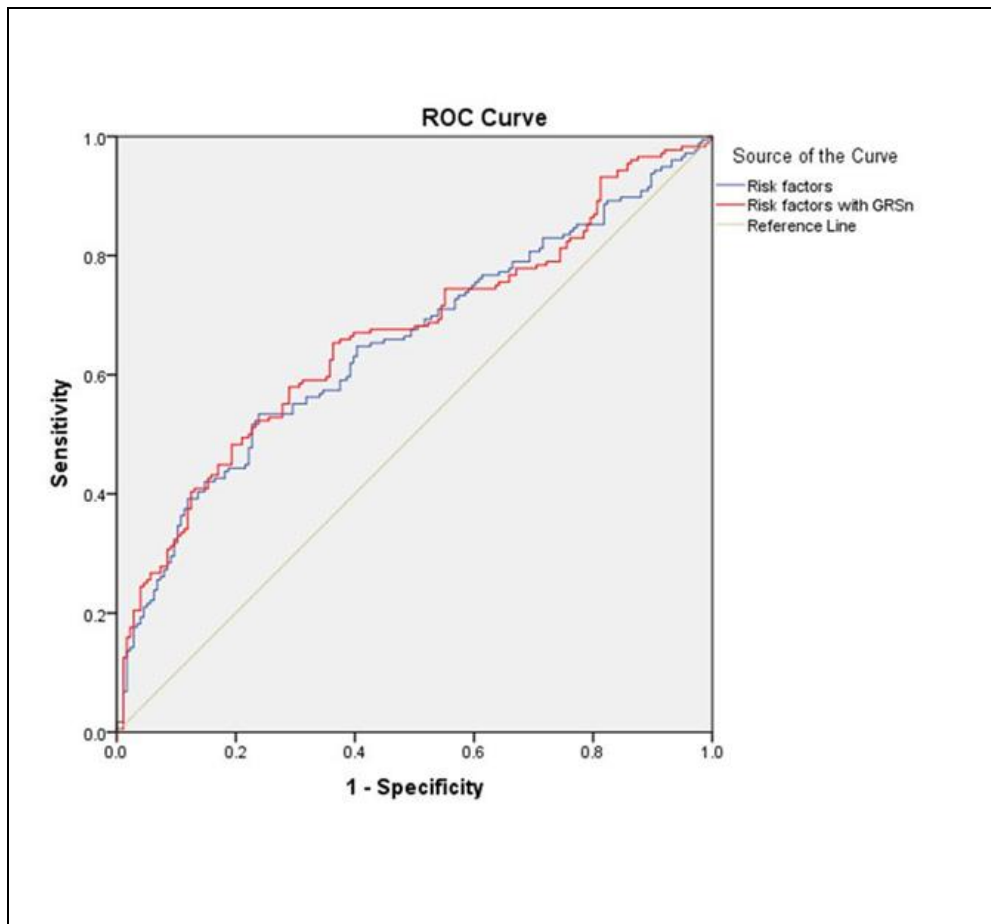


Figure 1 Receiver operating curve analysis for selected T2D risk factors. The T2D risk factors (age, urbanisation, BMI, fasting glucose and glycated hemoglobin (HbA1C) and GRSn) ability to predict cases was evaluated using ROC

4.0 Discussion

We set out to determine the GRS associated with increased T2D risk in the black South African population of Setswana descent. We found that the GRSn which comprised of variants that had significant associations in this population, to be significantly associated with increased T2D risk as indicated by an OR of 1.21 (1.02 - 1.43). The GRSn modestly

improved the predictive ability of the conventional T2D risk factors as had been reported for other GRS in earlier studies among the Europeans, Asians and African Americans [7, 9, 40]. The GRS was also noted to be more predictive of risk among the non-obese participants younger than 50 years of age. Thereby, indicating the attractiveness of using the GRS in early identification and prevention strategies.

The GRSn significantly associated with T2D as was expected, since the SNPs which it consisted of, had previously been reported to be associated with T2D in this population group. However, the GRSt consisting of the selected 66 variants in this study was noted to be significantly less predictive [OR 0.95 (0.90-1.00)] against T2D risk, contrary to what was noted in studies among Caucasians where GRSt of more than 60 variants were associated with increased T2D risk as depicted by OR's of 1.13 [41], 1.03 [42] and 1.08 [9]. Similarly, the GRSt was associated with increased risk (OR 1.05) among African Americans [9, 40]. The discrepancy between the results of the GRSt in our study and the other ethnicities might be due to the differences in linkage disequilibrium (LD) of the SNPs included in the GRS of the black South Africans with other ethnicities [11]. However, Vassy et al. [9] suggest that the polygenic risk score might be a proxy for the count of the loci which might be associated with T2D among different ethnicities such that the GRSt may be associated with T2D in multiple ethnicities. However, it's still unclear if the loci are associated with T2D across multiple ethnicities, since no overlap of the chromosomal regions associated with T2D in Caucasians and Africans were noted in linkage association studies, and the causal genes are still yet to be identified [43]. The advent of the customized GWAS genotyping chips in African populations will further help in

identifying the population specific variants and variants shared by Africans with other ethnicities associated with T2D as GWAS genotyping chips derived from European ethnicities are less effective in genetic screening of African populations [44, 45]. A recent review by Smith et al. [15], suggests that the GRS has to be population specific due to the differences in allele frequencies, LD and ethnicity-specific genetic variation that exist among the varied ethnic groups. Thus, indicating the need to identify population specific GRS for predicting T2D risk among the black South Africans.

It has been suggested that constructing a GRS based on the biological function of variants is more pivotal as compared to using only significant SNPs [9]. The GRS beta which was used in this study was however, associated with protection against T2D and not an increase T2D risk. It is most probable that the approach is valuable as has been documented in other studies, but that the variants associated with beta function in the black South African population might be different [9]. Therefore, it will be pivotal to explore this approach using population specific variants associated with beta function in the future.

There is an understanding that some of the variants associated with T2D are shared among different ethnicities as has been documented in trans-ethnic meta-analysis [46]. Therefore, we selected shared T2D associated variants to develop the GRStrans from one of the largest T2D trans-ethnic meta-analysis for T2D [47]. Nevertheless, the GRStrans was significantly less predictive against T2D risk in our study population. In view that the GRStrans, did not comprise of variants discovered in black South Africans,

differences in LD as noted for the GRSt might have led to it being less predictive [11, 15, 48].

The GRSn had significant interactions with Age and BMI in separately of P values equal to 0.033 and < 0.001 respectively. These interactions were further explored in Age and BMI stratified associations of the GRSn with T2D risk. The GRSn was noted to be more predictive to risk among less than 50 years old and non-obese participants compared to the obese and 50 years or more participants as indicated in Table 4. Though this finding among this population group maybe first being indicated in our study, it has been reported among the Asian and European population groups [8, 49].

In view that T2D is heritable, it is expected to depict the genetic risk early in life before the development of other risk factors such as obesity have been noted. The GRSn is a proxy of the T2D genetic risk and hence its improved predictiveness among the less 50 year old participants was expected , and other GRS proxies by de Miguel-Yanes et al [8] also reported similar results [8]. Evidence from a longitudinal study in which a GRS was noted to be more predictive of T2D risk in young adults aged 18-30 years for a period of for 25 years provides strong support of the GRS approach being more predictive in risk in people less than 50 years as depicted in our study [50].

With regards to the GRS interactions with obesity, Perry et al reported a per risk allele OR of 1.13 (P value = 2.2×10^{-14}) among the lean T2D cases as opposed to an OR of 1.06 (P value = 2.2×10^{-16}) in the obese among the Caucasians indicating the improved

predictiveness of the GRS among lean T2D patients [51]. Similarly this phenomena has recently been reported by Kong et al in a Han Chinese population [49]. The high heterogeneity of T2D is postulated to be responsible for the differences in genetic risks among the obese and non-obese cases. Epidemiological studies indicate that the non-obese patients develop rapid β cell failure in the condition of insulin sensitivity thereby requiring early insulin treatment compared to the obese [52]. This heterogeneity might be as result of different factors which include genetics being involved in the T2D etiology among obese and non-obese patients [53]. The majority of the known T2D variants are postulated to be related to beta cell function, thereby explaining the improved predictiveness of the GRSn in the lean patients who are have been noted to be more susceptible to β cell failure in epidemiological studies. However, since the function of the variants of the SNPs in our GRSn which was more predictive among the non-obese patients in our study is not known as they reside in intergenic regions there might be plausibly related to β cell function.

5.0 Conclusion

In view of our findings, the GRS approach is of less clinical use in the black South African population due to its modest improvement in predicting T2D risk. Since T2D, has a polygenic etiology, the GRS approach is being further limited by the few variants significantly associated with T2D in the black South African context. Thus, direct consumer genetic testing is of less clinical use in the black South African population as has been documented in European and Asian ethnicities [7, 10]. The advent of the Afrocentric customized genotyping array gives hope for the possibility of identifying more

population sensitive variants which might improve the utility of the GRS in the future. The short LD blocks among African populations, ensure an increased chance of identify the causal variants in future GWAS in Africa. These causal variants combined into a GRS might be more predictive of T2D risk [11]. The pathway based analysis of these population variants and inclusion of rare variants in the GRS might in the future yield a more clinically useful GRS for early identification of high risk T2D patients.

Conflict of Interest Statement

This research was supported by a grant National Research Foundation of South Africa (87970). Tinashe Chikowore is supported by the National Research Foundation of South Africa (89117). We would like to acknowledge Prof A Kruger and the PURE research team, funders and participants. All the authors have no conflicts of interest to declare.

References

- [1] Bradshaw D, Norman R, Pieterse D, Levitt NS, South African Comparative Risk Assessment Collaborating G. Estimating the burden of disease attributable to diabetes in South Africa in 2000. South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde. 2007;97:700-6.
- [2] Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes research and clinical practice. 2011;94:311-21.
- [3] Groop L, Pociot F. Genetics of diabetes—Are we missing the genes or the disease? Molecular and cellular endocrinology. 2014;382:726-39.
- [4] Saudek CD, Herman WH, Sacks DB, Bergenstal RM, Edelman D, Davidson MB. A new look at screening and diagnosing diabetes mellitus. The Journal of clinical endocrinology and metabolism. 2008;93:2447-53.
- [5] Florez JC. Chapter 82 - Diabetes. In: Ginsburg GS, Willard HF, editors. Genomic and Personalized Medicine (Second Edition): Academic Press; 2013. p. 990-1005.

- [6] Pyke DA. Development of diabetes in identical twins. *Adv Exp Med Biol*. 1988;246:255-8.
- [7] Wang X, Strizich G, Hu Y, Wang T, Kaplan RC, Qi Q. Genetic markers of type 2 diabetes: Progress in genome-wide association studies and clinical application for risk prediction. *Journal of diabetes*. 2016 8(1):24-35
- [8] de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, et al. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes care*. 2011;34:121-5.
- [9] Vassy JL, Hivert MF, Porneala B, Dauriz M, Florez JC, Dupuis J, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes*. 2014;63:2172-82.
- [10] Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisterman EF, et al. Predicting risk of type 2 diabetes mellitus with genetic risk models on the basis of established genome-wide association markers: a systematic review. *American journal of epidemiology*. 2013;178:1197-207.
- [11] Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nature reviews Genetics*. 2010;11:149-60.
- [12] Chikowore T, Conradie KR, Towers GW, van Zyl T. Common Variants Associated with Type 2 Diabetes in a Black South African Population of Setswana Descent: African Populations Diverge. *OMICS*. 2015;19:617-26.
- [13] Muller B, Wilcke A, Boulesteix AL, Brauer J, Passarge E, Boltze J, et al. Improved prediction of complex diseases by common genetic markers: state of the art and further perspectives. *Hum Genet*. 2016;135:259-72.
- [14] Chen R, Corona E, Sikora M, Dudley JT, Morgan AA, Moreno-Estrada A, et al. Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet*. 2012;8:e1002621.
- [15] Smith JA, Ware EB, Middha P, Beacher L, Kardia SL. Current Applications of Genetic Risk Scores to Cardiovascular Outcomes and Subclinical Phenotypes. *Current epidemiology reports*. 2015;2:180-90.
- [16] Adeyemo AA, Tekola-Ayele F, Doumatey AP, Bentley AR, Chen G, Huang H, et al. Evaluation of Genome Wide Association Study Associated Type 2 Diabetes Susceptibility Loci in Sub Saharan Africans. *Front Genet*. 2015;6:335.
- [17] Nelson SC, Doheny KF, Pugh EW, Romm JM, Ling H, Laurie CA, et al. Imputation-based genomic coverage assessments of current human genotyping arrays. *G3 (Bethesda)*. 2013;3:1795-807.

- [18] Morris AP. Fine mapping of type 2 diabetes susceptibility loci. *Curr Diab Rep*. 2014;14:549.
- [19] Fontaine-Bisson B, Renstrom F, Rolandsson O, Magic, Payne F, Hallmans G, et al. Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population. *Diabetologia*. 2010;53:2155-62.
- [20] Poveda A, Koivula RW, Ahmad S, Barroso I, Hallmans G, Johansson I, et al. Innate biology versus lifestyle behaviour in the aetiology of obesity and type 2 diabetes: the GLACIER Study. *Diabetologia*. 2016;59:462-71.
- [21] Qi Q, Hu FB. Genetics of type 2 diabetes in European populations. *Journal of diabetes*. 2012;4:203-12.
- [22] Thomsen SK, Gloyn AL. The pancreatic beta cell: recent insights from human genetics. *Trends Endocrinol Metab*. 2014;25:425-34.
- [23] Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*. 2012;44:981-90.
- [24] Sim X, Ong RT, Suo C, Tay WT, Liu J, Ng DP, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS genetics*. 2011;7:e1001363.
- [25] Zeggini E, Scott LJ, Saxena R, Voight BF. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature genetics*. 2008;40:638-45.
- [26] Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*. 2007;316:1341-5.
- [27] Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature genetics*. 2010;42:579-89.
- [28] Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proença C, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nature genetics*. 2009;41:1110-5.
- [29] Shu XO, Long J, Cai Q, Qi L, Xiang YB, Cho YS, et al. Identification of new genetic risk variants for type 2 diabetes. *PLoS genetics*. 2010;6:e1001127.
- [30] Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nature genetics*. 2011;43:984-9.

- [31] Cho YS, Chen CH, Hu C, Long J, Ong RT, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nature genetics*. 2011;44:67-72.
- [32] Tsai FJ, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS genetics*. 2010;6:e1000847.
- [33] Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, An SS, et al. A genome-wide association search for type 2 diabetes genes in African Americans. *PloS one*. 2012;7:e29202.
- [34] Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316:1336-41.
- [35] Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *European Journal of Human Genetics*. 2012;20:801-5.
- [36] Yamauchi T, Hara K, Maeda S, Yasuda K, Takahashi A, Horikoshi M, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nature genetics*. 2010;42:864-8.
- [37] Parra EJ, Below JE, Krithika S, Valladares A, Barta JL, Cox NJ, et al. 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*. 2011;54:2038-46.
- [38] Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of BioMedical R, Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316:1331-6.
- [39] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007;81:559-75.
- [40] Imamura M, Shigemizu D, Tsunoda T, Iwata M, Maegawa H, Watada H, et al. Assessing the clinical utility of a genetic risk score constructed using 49 susceptibility alleles for type 2 diabetes in a Japanese population. *J Clin Endocrinol Metab*. 2013;98:E1667-73.
- [41] Walford GA, Porneala BC, Dauriz M, Vassy JL, Cheng S, Rhee EP, et al. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. *Diabetes care*. 2014;37:2508-14.

- [42] Vaxillaire M, Yengo L, Lobbens S, Rocheleau G, Eury E, Lantieri O, et al. Type 2 diabetes-related genetic risk scores associated with variations in fasting plasma glucose and development of impaired glucose homeostasis in the prospective DESIR study. *Diabetologia*. 2014;57:1601-10.
- [43] Ng MCY. Genetics of Type 2 Diabetes in African Americans. *Curr Diabetes Rep*. 2015;15:74
- [44] Dandara C, Huzair F, Borda-Rodriguez A, Chirikure S, Okpechi I, Warnich L, et al. H3Africa and the African life sciences ecosystem: building sustainable innovation. *Omics : a journal of integrative biology*. 2014;18:733-9.
- [45] May A, Hazelhurst S, Li Y, Norris SA, Govind N, Tikly M, et al. Genetic diversity in black South Africans from Soweto. *BMC Genomics*. 2013;14:644.
- [46] Waters KM, Stram DO, Hassanein MT, Le Marchand L, Wilkens LR, Maskarinec G, et al. Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS genetics*. 2010;6:10..
- [47] Replication DG, Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes C, South Asian Type 2 Diabetes C, Mexican American Type 2 Diabetes C, Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples C, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics*. 2014;46:234-44.
- [48] Mhandire K, Pharo G, Kandawasvika GQ, Duri K, Swart M, Stray-Pedersen B, et al. How does mother-to-child transmission of HIV differ among African populations? Lessons from MBL2 genetic variation in Zimbabweans. *Omics : a journal of integrative biology*. 2014;18:454-60.
- [49] Kong X, Xing X, Hong J, Zhang X, Yang W. Genetic variants associated with lean and obese type 2 diabetes in a Han Chinese population: A case-control study. *Medicine (Baltimore)*. 2016;95:e3841.
- [50] Vassy JL, Durant NH, Kabagambe EK, Carnethon MR, Rasmussen-Torvik LJ, Fornage M, et al. A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia*. 2012;55:2604-12.
- [51] Perry JR, Voight BF, Yengo L, Amin N, Dupuis J, Ganster M, et al. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet*. 2012;8:e1002741.
- [52] George AM, Jacob AG, Fogelfeld L. Lean diabetes mellitus: An emerging entity in the era of obesity. *World J Diabetes*. 2015;6:613-20.

[53] Timpson NJ, Lindgren CM, Weedon MN, Randall J, Ouwehand WH, Strachan DP, et al. Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes*. 2009;58:505-10.

CHAPTER 4

Exploration of the genetic variants associated with type 2 diabetes using novel whole genome sequences in individuals of Setswana descent in South Africa

Exploration of the genetic variants associated with type 2 diabetes using novel whole-genome sequences in individuals of Setswana descent in South Africa

Tinashe Chikowore^{2*}, Shaun Aron³, Andrew P. Morris⁴, Scott Hazelhurst³, Tertia van Zyl¹, E Feskens², Michele Ramsay³ and Karin R. Conradie¹

¹Centre for Excellence in Nutrition, North-West University, Potchefstroom, North West Province, 2520, South Africa;

² Wageningen University, Division of Human Nutrition, P.O. Box 17, 6700 AA Wageningen, The Netherlands

³ Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 9 Jubilee Road, Parktown, 2050

⁴ Department of Biostatistics, University of Liverpool, Liverpool, UK

* Corresponding author

E-mail: tinashedoc@gmail.com;

Abstract

The burden of Type 2 diabetes is a major global public health concern. In Africa, where it is estimated that T2D prevalence will have the highest increase by 2030, more efforts are now required to curb this burden. In view of the strong heritability of T2D, unravelling the underlying genetic factors might help in the development of effective and personalised interventions to limit this disease. The genetic determinants of T2D, however, have not been explored using whole-genome sequences that can enable the detection of rare variants which are postulated to have large effects on T2D risk in Africa. We therefore evaluated the genetic variants associated with T2D among 16 T2D cases and 14 controls that were selected from the extremes of the glycated hemoglobin distribution. Single variant and gene-based tests were performed while adjusting for age, sex, and BMI and population structure using linear mixed modelling. The rs7499753 SNP, located in the intron of *CIITA*, attained association with T2D risk beyond the genome-wide significance level at $p = 1.45 \times 10^{-8}$. Some coding SNPs with nominal significance ($p < 5 \times 10^{-5}$), which will need to be validated in future studies, were discovered. Through collaborative efforts, we noted that the Setswana, Zulu and Sotho ethno-linguistic groups shared a common ancestry and might be closely related. However, well-powered collaborative whole genome sequencing initiatives are required to elucidate the genetic variants associated with T2D among black South Africans.

Keywords: Type 2 diabetes, genetics, SNPs, ethno-linguistic groups, black South Africans

Introduction

Type 2 diabetes (T2D) is a cause of concern in global public health. It is postulated that, by 2030, the largest increase of 109% in the prevalence of T2D will occur in Africa (1). Sadly, most African countries are ill-equipped to manage the current T2D burden and the possibility of future increases of this disease will further exhaust resources and could lead to more loss of lives if drastic measures are not taken to curtail this pandemic (2). T2D poses a further threat to economic growth and development, as it has been noted to be a significant cause of disability and mortality in the workforce of productive age in South Africa (3). Urgent strategies are therefore required to curb this increase.

Efforts to control this pandemic are being held back by lack of clarity about the precise factors involved in the aetiology of T2D (4). Of particular concern is the failure to determine genetic factors that account for the high heritability of T2D of up to 50% (5). To date, about 135 genetic variants have been associated with T2D in European and Asian ethnicities through the Genome Wide Association Study (GWAS) approach (6). However, the majority of these SNPs have been noted to be of less clinical utility as they explain about 10% of the heritability of T2D among Europeans and have modest effects on T2D risk (7). It is postulated that a proportion of the missing heritability can be accounted for by rare variants which are not well captured by the GWAS approach (8).

The extreme phenotype sampling approach in sequence-based studies is among the attractive methods that can be used to detect rare variants associations with T2D (9). In the extreme phenotype sampling approach, participants are selected from the less than

10% tails of a distribution curve of standardised residual probabilities after the adjustment of covariates for a quantitative trait (10). This enables the participants to be sampled from the tails of the unexplained variation of the quantitative trait; these are most likely to be accounted for by genetic factors as they would not be explained by known covariates (10). Although reductions in sequencing costs have occurred in recent years, many samples are still required to conduct rare variant association tests using random sampling approaches, which will in the end make such a project financially inaccessible (9). However, the extreme phenotype sampling strategy which enriches rare variants in the extremes has been shown to improve power in identifying causal variants as opposed to random sampling, while using smaller sample sizes from the study population (11). Using extreme phenotype sampling, the nonsense variant in *SLC30A8* was identified and, after replication, was reported to reduce T2D risk by 53% (12). In contrast, a recent trans-ethnic study among 44 414 individuals suggested that low-frequency and rare variants might not have the large effects for T2D risk that they are postulated to have (13). However, the role of the rare and low frequency variants with regard to T2D risk among the African population residing in Africa has yet to be determined.

Africa has been lagging behind in deciphering the genetic factors associated with T2D susceptibility while, this region, alarmingly, is postulated to have the highest increase in this disease (1). Of particular concern is South Africa, which has the highest reported age-standardised prevalence of T2D in sub-Saharan Africa of 13.1% among urban adults in Cape Town (14). Although GWAS of T2D have been performed among African Americans of mixed ancestry, no GWAS for T2D has been performed in Africa (15). However,

recently the whole-genome sequences of the black Zulu and Sotho South African ethno-linguistic groups were sequenced in the African Genome Variation Project (AGVP) (16). In South Africa, there are about eleven of these ethno-linguistic groups among black South Africans and it is not known how they are related to each other in view of the great genetic diversity that exists in African populations. Efforts to unravel the genetic relatedness of these ethno-linguistic groups are pivotal as they will enable the application of future genetic initiatives to be specified according to the genetic similarities and differences among these groups to curb T2D. This study, therefore, seeks to 1) determine the genetic relatedness of the Setswana and other Zulu, Sotho ethno-linguistic groups; 2) to identify variants associated with T2D among the Setswana population group, for targeted follow-up validation studies.

Results

Descriptive characteristics of study participants

The study participants that were sampled from the upper extreme had almost significantly double the glycated haemoglobin and fasting glucose levels of those sampled from the lower extreme. Systolic and diastolic blood pressures were also significantly higher in the upper extreme group compared with the lower extreme group, as indicated in Table 1.

Table 1 Descriptive characteristics of the study participants

Variable	Case (upper extreme)	Controls (lower extreme)	P value
n	16	14	
Gender n (males/females)	3/13	5/9	0.263
Age (2005)	58.38 ± 11.28	53.50 ± 9.96	0.223
BMI (kg/m ²)	32.61 ± 7.84	28.46 ± 8.53	0.172
Fasting glucose (mmol/l)	11.49 ± 8.18	5.34 ± 1.25	0.010
Glycated haemoglobin (%)	9.46 ± 3.25	4.98 ± 0.61	< 0.001
Systolic blood pressure (mmHg)	153.63 ± 22.26	127.29 ± 16.18	0.001
Diastolic blood pressure (mmHg)	98.13 ± 8.42	84.71 ± 12.43	0.002

BMI=body mass index

Variants from whole genome sequence and QC

About 4.7 million variants were called per individual and 200 000 of these were singletons. Across all individuals, 54% of the variants were novel according to the dbSNP 131. Overall, 19 162 050 SNPs were called for all the 30 samples. However, 18 563 881 bi-allelic SNPs only were prioritised for QC. 18 251 063 SNPs were retained after the QC and used for downstream analysis.

Principal component analysis and genetic relatedness

The Setswana population group, which was sequenced in this study, was noted to share a common ancestry with the other ethno-linguistic Zulu and Sotho groups from South Africa that evaluated in the AGVP; and thus probably closely related to them, based on principle component analysis (PCA) of 342 755 independent SNPs.(Figure S1 and S2)

Single variant tests of association with T2D

A total of 10 488 293 SNPs (MAF > 1%) were evaluated in the single variant tests using the EMMAX linear mixed model and only one variant rs7499753 located in the intron of *CIITA* gene was noted to exceed the genome-wide significance level with a p value of 1.45×10^{-8} . Notably there were some nominally significant ($p < 5 \times 10^{-5}$) coding SNPs, as illustrated in Table 2.

Table 2 Selected SNPs associated with T2D through the single variant test approach

Chr	SNP	Alleles	MAF	Annotation:Gene	Case Geno Count	Cont Geno Count	OR (95CI)	P value
16	rs7499753	G/A	0.28	Intron: <i>CIITA</i>	03/10/03	13/1/0	9.65 (5.62-16.58)	1.45×10^{-8}
11	rs73543086	A/C	0.22	Intergenic	04/12/00	13/1/0	19.45 (8.82-42.89)	1.05×10^{-7}
18	rs7239485	G/A	0.40	Intron: <i>ZNF532</i>	04/06/06	09/04/01	7.95 (4.42- 14.31)	3×10^{-7}
1	rs2796378	G/C	0.45	Intergenic	08/08/00	0/9/5	0.11 (0.06-0.20)	3.34×10^{-7}
3	rs9867221	G/A	0.38	Intron: <i>RP11-644C3.1</i>	09/07/00	01/10/03	0.09 (0.05-0.18)	6.02×10^{-7}
7	rs328909	A/G	0.20	Exon: <i>DPY19L1</i>	15/1/0	04/09/01	0.07 (0.03-0.19)	1.94×10^{-5}
1	rs896318	A/G	0.23	Exon: <i>RP11-383G10.5</i>	06/07/03	13/1/0	6.85 (3.34-14.08)	2.00×10^{-5}
1	rs3863779	C/T	0.37	Exon: <i>NBPF1</i>	07/09/00	01/13/00	0.05 (0.01-0.15)	3.47×10^{-5}
14	rs8015899	A/G	0.35	Exon: <i>CTD-2506J14.1</i>	02/11/03	10/04/00	8.57 (3.66-20.07)	4.30×10^{-5}

Chr=chromosome; Geno=genotype; Cont=control; MAF = minor allele frequency; OR=odds ratio

Gene-based tests and rare-variant associations with T2D

A total of 16 114 genes with rare nonsynonymous variants (MAF 0.1% - 5%) were included in the gene-based tests, using the SKAT-O test with EMMAX modelling for adjusting for population structure. None of the genes with rare variants was able to reach genome wide gene-based significance (p value $< 2.8 \times 10^{-6}$), as indicated in Figure S5 and S6.

Discussion

We set out to determine the genetic relatedness of the Setswana, Zulu and Sotho ethno-linguistic groups that reside in South Africa and to discover the variants associated with T2D among the novel whole-genome sequences in the black South African population of Setswana descent. We sequenced, for the first time, the whole genomes of black South Africans of Setswana descent. About 54% of the variants called from the whole genomes sequences were novel according to the dbSNP 131 version that was used to annotate these variants. We noted that the Setswana, Zulu and Sotho people, although they speak different languages, are genetically closely related, as indicated in Figure S5 and S6. We discovered an association of rs7499753 located in the intron of *CIITA* with T2D risk beyond the GWAS significance level at p value = 1.45×10^{-8} that's pending validation in replication studies. This was the only SNP which was significantly associated with T2D, as illustrated in Figure S3 and S4, during the single variant association tests. Although it is postulated that the rare variants are enriched in the tails of extreme phenotype sampling, there was no gene with rare variants that surpassed the genome-wide significance level, as illustrated in Figure S5 and S6.

To the best of our knowledge, our study is the first to sequence the whole genomes of black South Africans of Setswana descent. Fifty four percent of the variants were novel, based on the dbSNP version 131. However, there is a need to evaluate whether some of the novel variants have not been captured in the latest dbSNP 149 version. Nonetheless, there is still a considerable amount of genetic variation in Africans which is yet to be captured. This further emphasises the need for custom SNP genotyping arrays for assessing the genetic variation in African populations. The African custom chip being developed by the Human Heredity and Health in Africa (H3A) consortium, based on whole-genome sequences, thus holds more promise of leading to the determination of variants that are of importance to Africans (17). In South Africa there are eleven ethno-linguistic population groups whose genetic relatedness is not known. We report the shared common ancestry of Zulu, Sotho and Setswana black South African ethno-linguistic groups, based on population structure analysis through PCA, illustrated in Figure S5 and S6. This realisation implies that these three groups might be closely related. Therefore, the determination of the disease-associated genetic variants among these ethno-linguistic groups will be enhanced, because the variants in one of these population groups might also be relevant in the others, thereby enabling a generalisation of prevention and treatment initiatives among these population groups.

The top association signal, rs7499753 in our study was reported not to be significant ($p = 0.324$) in a recent a study by Fuchsberger et al (13), and it was indicated to have a MAF of 0.068 in the Europeans, which was lower than the MAF of 0.28 we noted in the South African population. Our MAF for the rs7499753 SNP was comparable to the MAF

in the 1000 Genomes Phase 3, dbSNP 149 for the Luhya African population, which is 0.21. Differences in allele frequencies and LD patterns have been noted to explain the varied associations of different SNPs in different ethnic groups (18-20). It is possible that rs7499753 is a population specific variant associated with T2D in our study ethnic group. However, further, well powered replication studies are required before the association of rs7499753 with T2D can be confirmed.

A number of coding SNPs had nominal significance levels ($p < 5 \times 10^{-5}$), as illustrated in Table 2. Most SNPs that have been discovered through the GWAS studies have been observed to be located in the introns (21). However, these exon-based variants need first to be validated in other independent studies, followed by functional studies to elucidate the mechanisms by which they affect T2D risk. These variants, which had nominal significant associations with T2D, together with our top signal variant, had higher ORs ranging from 0.05 to 19.45. In view, of the small sample sizes of our study, the SNPs with greater effects with T2D are mostly likely to yield significant associations. However, it is possible that these ORs might be inflated as logistic regression is reported to yield inflated ORs in small samples sizes as in our study (22, 23). Future replication studies that are well powered will thus be required to obtain more reliable estimates of the effect sizes of the discovered variants.

Although the selection of individuals sampled from the extremes of the cohort sample size is postulated to improve the ability to discover the rare variants, we were not able to detect significant rare-variant associations with T2D in our study. However, a previous larger

study of 758 participants was initially not able to detect genes with rare variants associated with T2D and had to include a second stage of replicating initially discovered variants to detect significant associations (12). Thus the use of two-stage extreme phenotype sampling might be pivotal in determining the rare variants associated with T2D in our setting, as suggested by Kang. (11). He indicates that, with as little as 0.2% of the study population in the first stage of extreme phenotype sampling of individuals from the 97th to 99th percentiles, about 80% of the causal variants may be discovered which will be detected in the second stage of replication (11). It is thus crucial to adopt and explore such a strategy in the future.

Our current study has some notable limitations, including having a modest sample size owing to limited resources. The results of this study, however, might later on be validated in a well-powered study and lead to detection of causal variants, as is claimed by Kang (11), through a two-stage extreme phenotype study design (11). The whole genomes sequences can be used in the future as imputation panels for GWAS South African studies. Nevertheless, our explorative study has pioneered the exploration of the genetic variants associated with T2D among black African populations using whole-genome sequencing data. This will help guide future research initiatives in this regard.

In conclusion, the genetic variants associated with T2D among black South Africans have yet to be clarified. However, we discovered that the rs7499753 based in the intron of *CIITA* was associated with T2D risk beyond the genome wide significance level, although this association needs to be validated in future replication studies. We hope that, with the

use of two-stage extreme phenotype sampling designs, and whole genome sequencing collaborations, more genetic variants, especially rare variants associated with T2D risk among black South Africans, will be determined.

Methods

Study population

Thirty male and female participants of Setswana descent were enrolled in the study. Sixteen T2D cases and 14 controls were selected from the 1% tails of the normal distribution curve of the standardised residuals of the multivariate linear model of glycated haemoglobin after adjustment for age, BMI, fasting glucose and gender among 2010 participants of the South African arm of the Prospective Urban, Rural and Epidemiological (PURE) Study. The research participants gave written informed consent before participating in the study. Ethical approval was granted by the Ethics Committee of the North-West University, Potchefstroom Campus, with ethics number NWU-00016-10-A1.

Whole-genome sequencing and quality control

The genomic DNA was extracted and normalised in the Nutrigenetics lab at the North-West University. The Qiagen DNA catalogue (51206 catalogue number, QIAGEN Pty. Ltd., Australia) extraction kit was used for the DNA extractions and the NanoDrop was then used to normalise the DNA to the required concentration. The Garvan Institute of Medical Research in Australia performed the whole-genome sequencing. The Truseq Nano and the PCR-Free library preparation kits were used to generate the libraries as per the manufacturer's instructions. One sample was loaded per flow cell and 2 x 150bp

sequencing at 30X coverage was performed on the Illumina HiSeq X sequencer (Illumina Inc., San Diego, CA, USA). The raw data were converted to FastQ file format using Illumina's bcl2fastq 2.16.0 (Illumina Inc., San Diego, CA, USA). The quality of the generated FastQ files was assessed using FastQC. The sequences were then aligned to the human reference genome (human.g1k.37) using iSAAC aligner v01.14.11.11 to generate BAM files, and additional quality metrics were evaluated using Picard WgsMetrics v1.119 (24). The GATK best practices pipeline was used to call and annotate the variants.

SNP and participant quality control

Quality control (QC) metrics for the participants and the bi-allelic SNPs were evaluated using PLINK (25). To control for missingness a genotype call rate of 5% was used as the cut for all the samples. The relatedness and common ancestry of the participants were evaluated using identity by descent metric (IBD) and the PCA respectively. Closely related individuals were indicated by an $IBD > 0.185$ and duplicates were noted by an $IBD > 0.98$. The clustering of the individuals on the PCA plots was used as a measure of common ancestry. The participants which did not cluster well with others were regarded as having a different ancestry from the other samples. The participants with outlying heterozygosity were removed based on a threshold cut off, of mean heterozygosity of all the samples ± 3 standard deviation. For SNPs the following QC metrics were evaluated, differential missingness between cases and controls, high missingness and Hardy Weinberg equilibrium (HWE). The SNPs which did not meet the expected cut-offs were

removed. That is for HWE ($p < 0.00001$), SNP missingness of (5%) and differential missingness ($p < 0.00001$).

PCA analysis of the Setswana population and the African Genome Variation Project data

The SNPs common between the Setswana sequenced samples from this studies and the Zulu and Sotho genotype data from the AVGP were merged in PLINK. The merged data set comprised of 1 026 901 common variants ($MAF > 5\%$). The SNPs in high LD were pruned out and the PCA was then performed on 342 755 independent SNPs. The Genesis software was used for the visualisation of the PCA plots.

Single variant and gene-based association tests

The EPACTS software was used to perform the association tests. In view of the presence of the population structure that was noted, the EMMAX linear mixed model that adjusts for hidden relatedness and population structure was used for single variant tests (26). Initially the kinship matrix for adjusting the population structure was created using variants with minimum MAF of 0.01 and call rate of 0.95. The EMMAX test in the EPACTS pipeline was used to perform the association tests while adjusting for age, sex and BMI. The EMMAX test is designed for quantitative traits as it yields a beta value and does not produce ORs. However, the p-values from the EMMAX test are applicable in binary outcomes like those of our study (26). We estimated the OR and its 95% confidence interval using the beta from the linear mixed model output of EMMAX together with its

standard error (sebeta) and a the proportion of the cases in the study as described elsewhere (27)

The gene-based association tests were done by first creating a gene-level grouping file for non-synonymous SNPs only. The association tests were conducted using the SKAT-O which accommodates EMMAX, while adjusting for age, sex, BMI using a maximum MAF of 5%. Quantile-quantile (QQ) plots and Manhattan plots were generated for the visualisation of the single variant and gene-based tests by the EPACTs software

Acknowledgements

We would like to acknowledge Prof. A Kruger and the PURE research team and participants. This study was supported by the National Research Foundation (NRF) of South Africa, with grant number 87970. None of the authors have any conflict of interest to declare. Andrew P Morris is a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant number WT098017).

References

1. Collaboration NCDRF. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet*. 2016;387(10027):1513-30.
2. Mbanya JC, Motala AA, Sobngwi E, Assah FK, Enoru ST. Diabetes in sub-Saharan Africa. *Lancet*. 2010;375(9733):2254-66.
3. Bradshaw D, Norman R, Pieterse D, Levitt NS, South African Comparative Risk Assessment Collaborating G. Estimating the burden of disease attributable to diabetes in

South Africa in 2000. South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde. 2007;97(8 Pt 2):700-6.

4. Nolan CJ, Damm P, Prentki M. Type 2 diabetes across generations: from pathophysiology to prevention and management. Lancet. 2011;378(9786):169-81.

5. Risch N. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. American journal of human genetics. 1990;46(2):242-53.

6. Prasad RB, Groop L. Genetics of type 2 diabetes-pitfalls and possibilities. Genes (Basel). 2015;6(1):87-123.

7. Qi Q, Hu FB. Genetics of type 2 diabetes in European populations. Journal of diabetes. 2012;4(3):203-12.

8. Sanghera DK, Blackett PR. Type 2 Diabetes Genetics: Beyond GWAS. Journal of diabetes & metabolism. 2012;3(198):6948.

9. Guey LT, Kravic J, Melander O, Burt NP, Laramie JM, Lyssenko V, et al. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. Genet Epidemiol. 2011;35(4):236-46.

10. Lanktree MB, Hegele RA, Schork NJ, Spence JD. Extremes of unexplained variation as a phenotype: an efficient approach for genome-wide association studies of cardiovascular disease. Circulation Cardiovascular genetics. 2010;3(2):215-21.

11. Kang G, Lin D, Hakonarson H, Chen J. Two-stage extreme phenotype sequencing design for discovering and testing common and rare genetic variants: efficiency and power. Human heredity. 2012;73(3):139-47.

12. Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burt NP, et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nature genetics. 2014;46(4):357-63.

13. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536(7614):41-+.

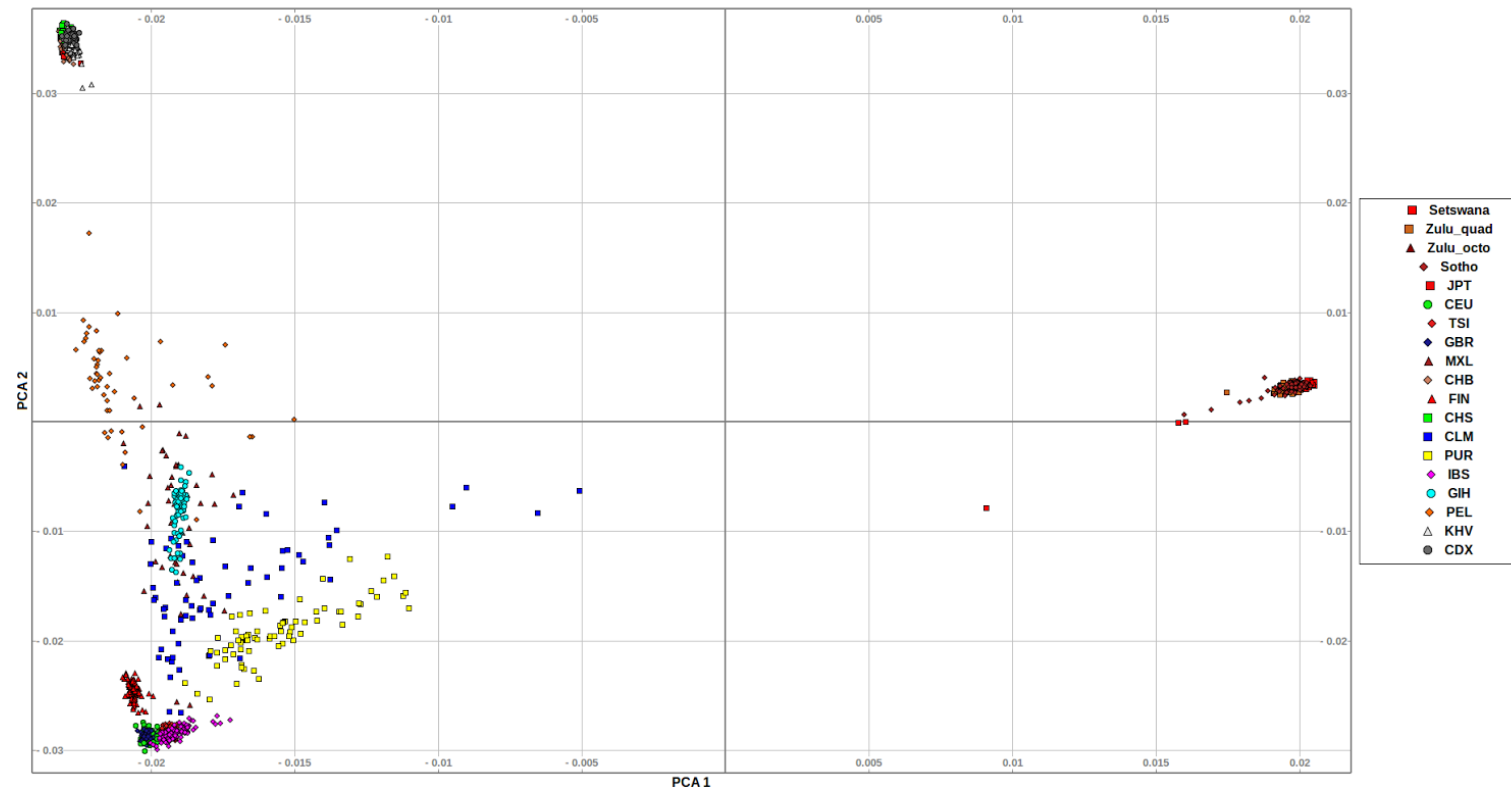
14. Peer N, Steyn K, Lombard C, Lambert EV, Vythilingum B, Levitt NS. Rising diabetes prevalence among urban-dwelling black South Africans. PloS one. 2012;7(9):e43336.

15. Ng MCY. Genetics of Type 2 Diabetes in African Americans. Curr Diabetes Rep. 2015;15(10).

16. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. Nature. 2015;517(7534):327-32.

17. Ramsay M, Sankoh O, as members of the AWIGs, the HAC. African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent. *Int J Epidemiol*. 2016;45(2):305-8.
18. Chen R, Corona E, Sikora M, Dudley JT, Morgan AA, Moreno-Estrada A, et al. Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet*. 2012;8(4):e1002621.
19. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nature reviews Genetics*. 2010;11(2):149-60.
20. Adeyemo AA, Tekola-Ayele F, Doumatey AP, Bentley AR, Chen G, Huang H, et al. Evaluation of Genome Wide Association Study Associated Type 2 Diabetes Susceptibility Loci in Sub Saharan Africans. *Front Genet*. 2015;6:335.
21. Groop L, Pociot F. Genetics of diabetes—Are we missing the genes or the disease? *Molecular and cellular endocrinology*. 2014;382(1):726-39.
22. Nemes S, Jonasson JM, Genell A, Steineck G. Bias in odds ratios by logistic regression modelling and sample size. *BMC Medical Research Methodology*. 2009;9(1):56.
23. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365-76.
24. Racz C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics (Oxford, England)*. 2013;29(16):2041-3.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007;81(3):559-75.
26. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*. 2010;42(4):348-54.
27. Cook JP, Mahajan A, Morris AP. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet*. 2016.

SUPPLEMENTARY SECTION



GBR=British in England and Scotland; CDX=Chinese Dai in Xishuangbanna, China; CEU=Utah residents with Northern and Western European ancestry; CHB=Han Chinese in Beijing, China; CHS=Southern Han Chinese; CLM=Colombians from Medellin, Colombia; FIN=Finnish in Finland; GIH=Gujarati Indian from Houston, Texas, USA; IBS=Iberian population in Spain; JPT=Japanese in Tokyo, Japan; KHV= Kinh in Ho Chi Minh City, Vietnam; MXL= Mexican ancestry from Los Angeles, USA; PUR, Puerto Ricans from Puerto Rico, TSI=Toscans in Italy

Figure S1 PCA for Setswana population and other world population groups based on 325 062 SNPs after removal of SNPs in high LD

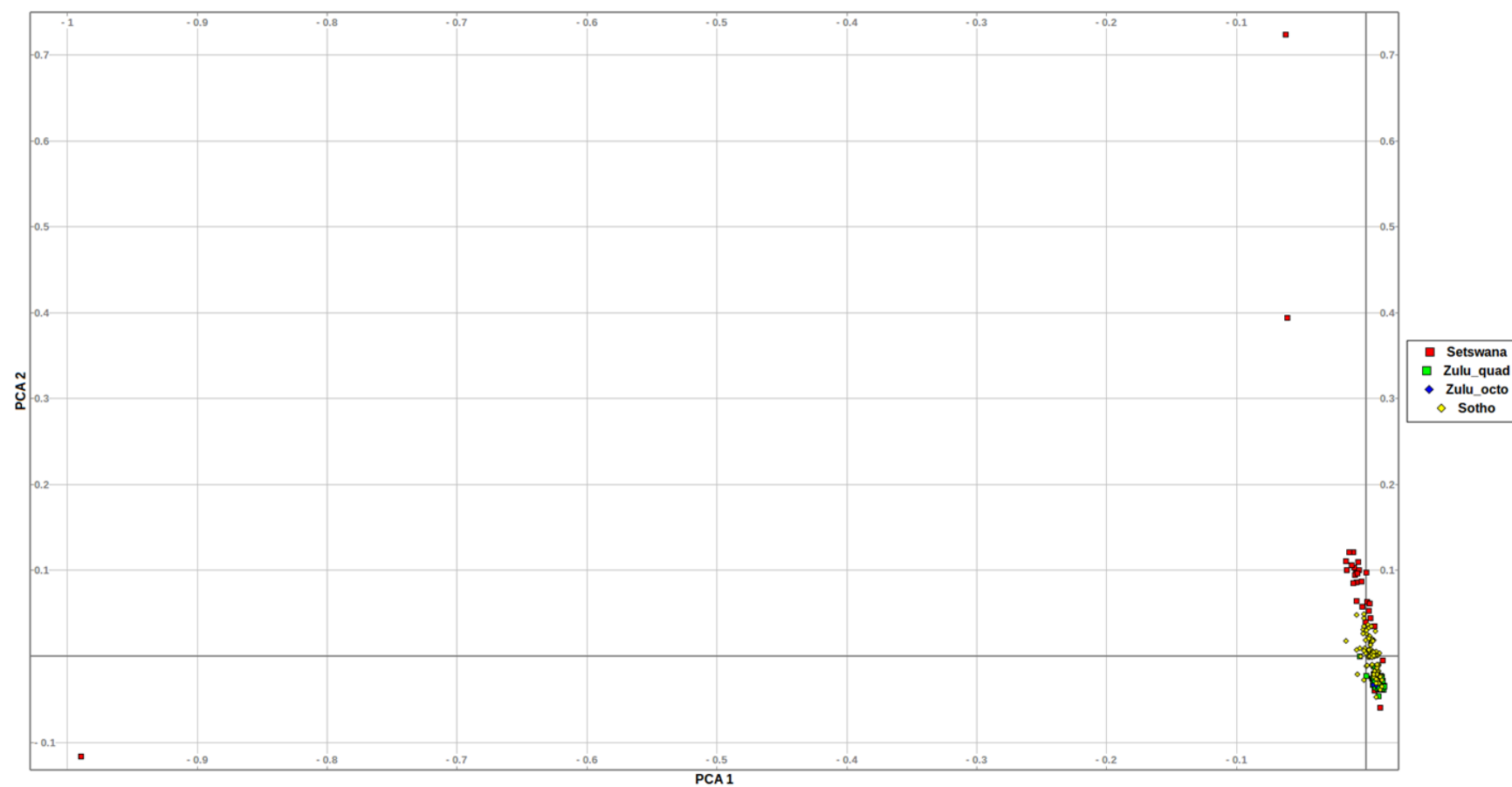


Figure S2 PCA of the of Zulu, Sotho and Setswana populations based on 342 755 SNPs after removal of SNPs in high LD

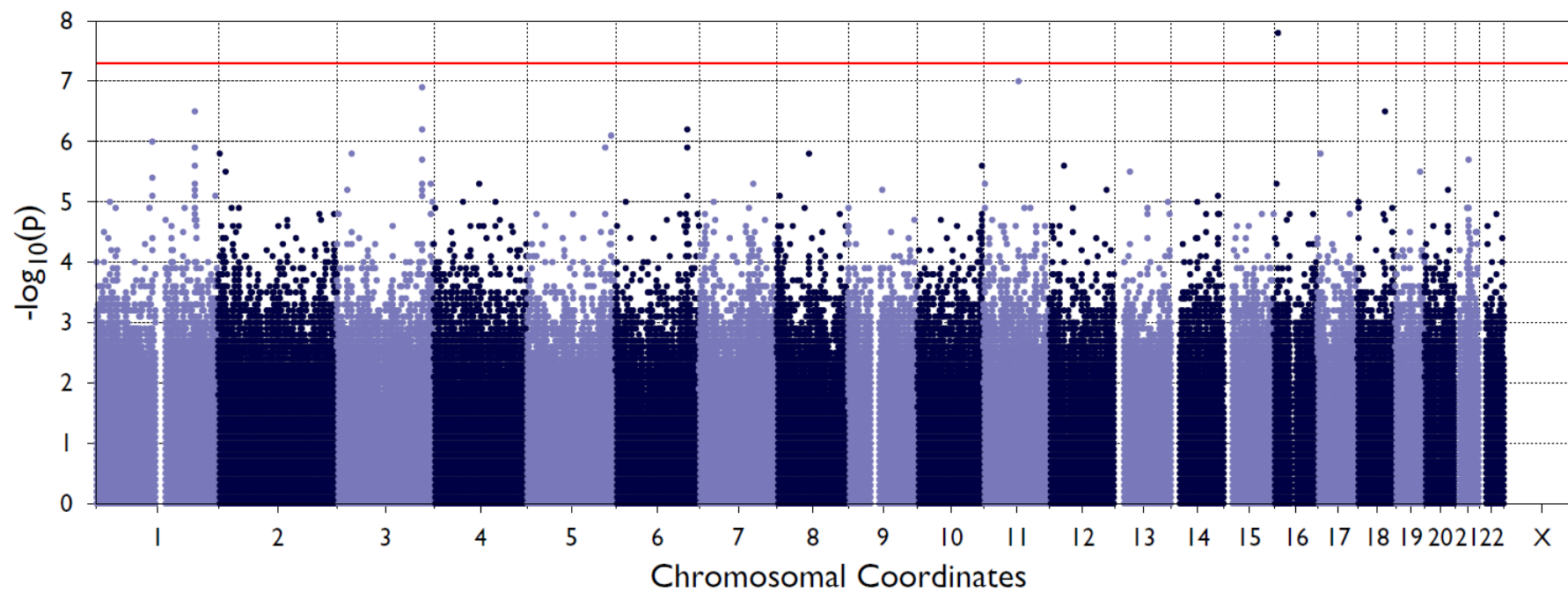


Figure S3 Manhattan plots for these single variant association results based on EMMAX and min MAF 0.01

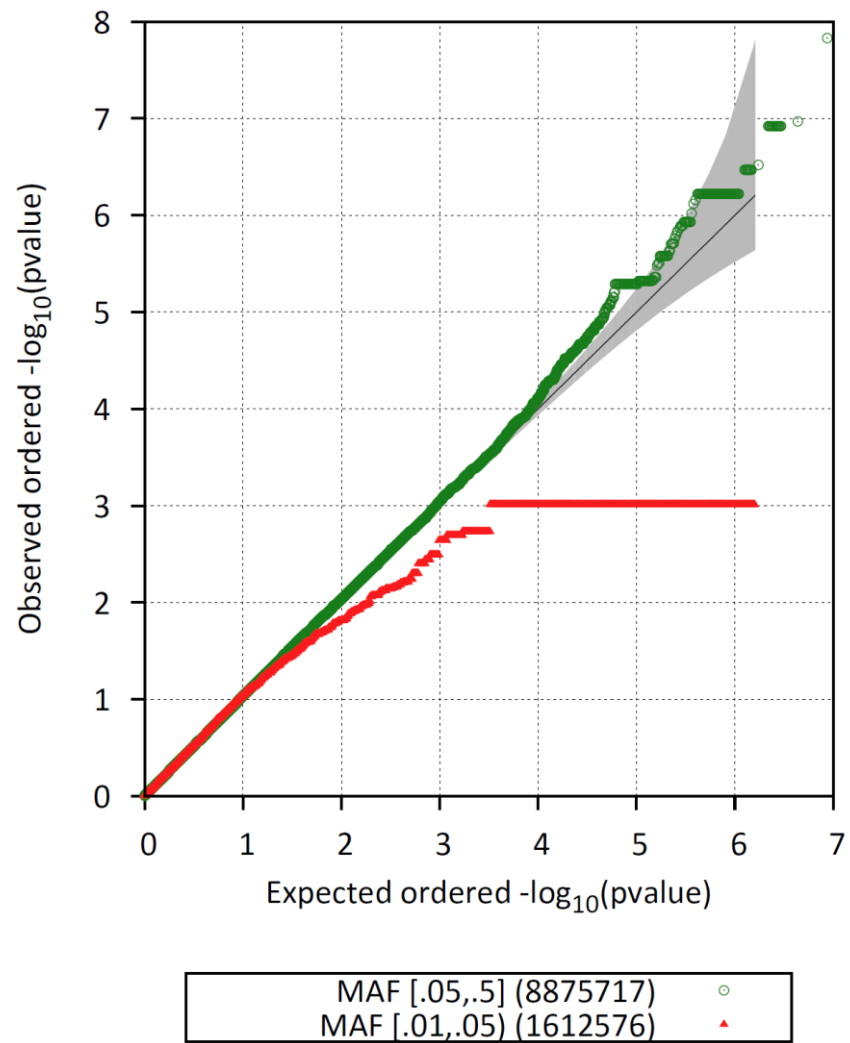


Figure S4 Q-Q plots for the single variant association tests using EMMAX, min MAF 0.01

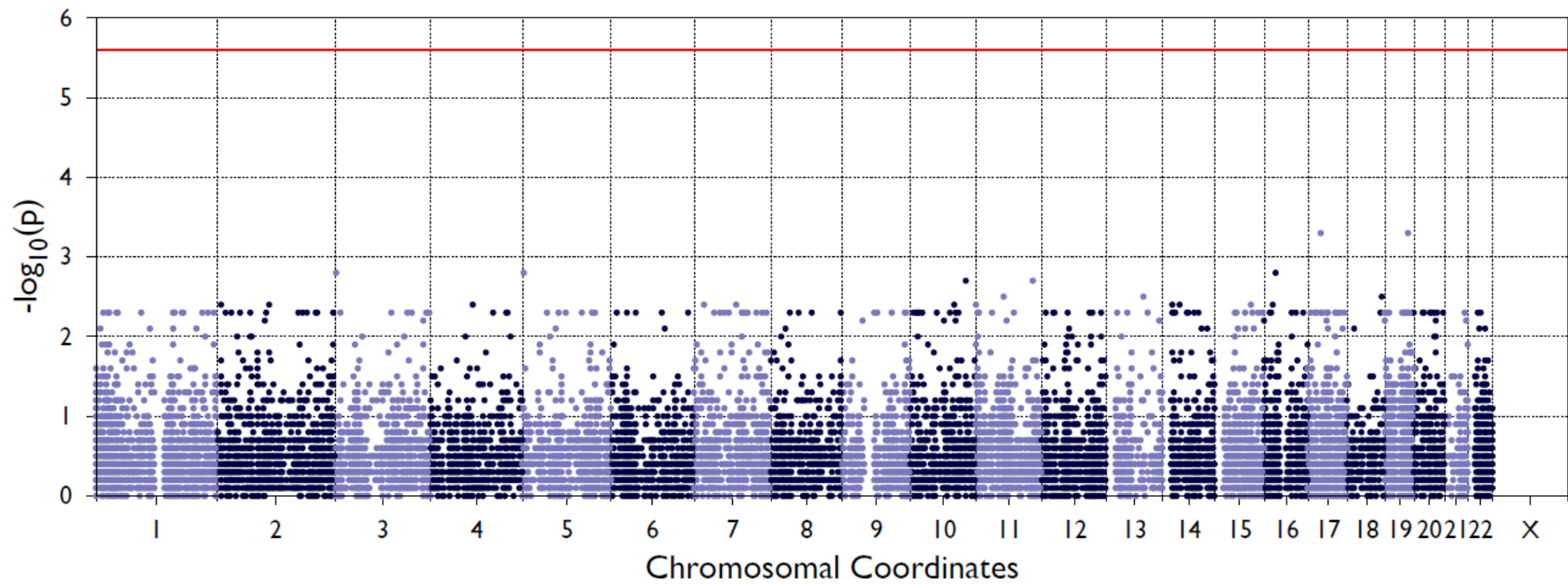


Figure S5 Manhattan plots for the gene-based tests using SKAT-O with EMMAX based on 16114 genes with MAF 0.1% to 5% of nonsynonymous variants

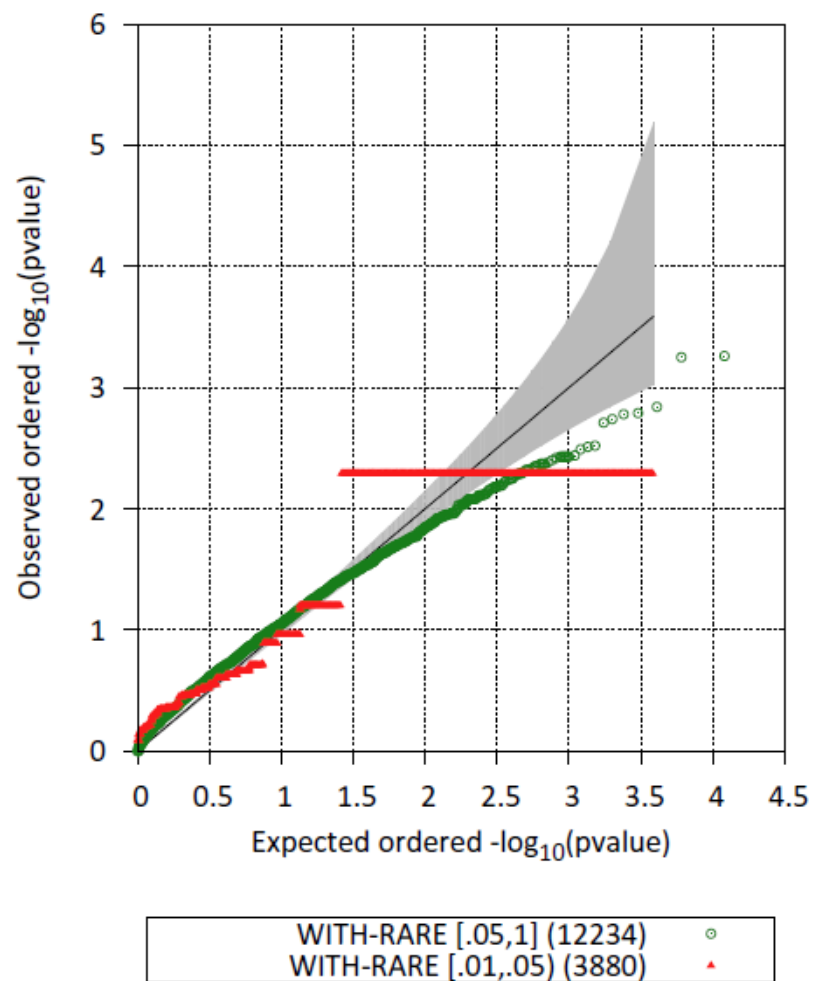


Figure S6 Q-Q plots for the gene-based SKAT-O test with EMMAX based on 16114 genes with MAF 0.1% to 5% of nonsynonymous variants

CHAPTER 5

Nutrient patterns associated with fasting and glycated hemoglobin in a black South African population

Nutrient patterns associated with fasting glucose and glycated hemoglobin levels in a black South African population

Tinashe Chikowore ^{1*}, Pedro T Pisa ², Tertia van Zyl ¹, Edith J.M. Feskens³, Edelweiss Wentzel-Viljoen ^{1,4} and Karin R Conradie ¹,

¹Centre for Excellence in Nutrition, North-West University, Potchefstroom, North West Province, 2520, South Africa

²Wits Reproductive Health and HIV Institute, University of the Witwatersrand, Johannesburg, South Africa

³ Wageningen University, Division of Human Nutrition, P.O. Box 17, 6700 AA Wageningen, The Netherlands

⁴ Medical Research Council Research Unit for Hypertension and Cardiovascular Disease, Faculty of Health Sciences, North-West University, Potchefstroom, South Africa

*Correspondence: tinashedoc@gmail.com ; Tel.: +27781169789

Abstract: Type 2 diabetes (T2D) burden is increasing globally. However, evidence regarding the nutrient patterns which are universal and directly involved in the aetiology of T2D than dietary patterns is scarce. This study set out to determine the nutrient patterns associated with fasting glucose and glycated hemoglobin. Factor analysis was used to derive nutrient patterns of 2010 participants stratified by urban/rural status and gender. Principal Component Analysis (PCA) was applied to 25 nutrients, computed from the quantified food frequency questionnaires (QFFQ). Three nutrient patterns per stratum which accounted for 73% of the variation of the selected nutrients were identified. Multivariate linear regression models adjusted for age, BMI, smoking, physical activity, education attained, alcohol intake, seasonality and total energy intake were computed. Starch, dietary fibre and B vitamins driven nutrient pattern was significantly associated with fasting glucose ($\beta = -0.236$ (-0.458; -0.014); $p = 0.037$), and glycated hemoglobin levels ($\beta = -0.175$ (-0.303; -0.047); $p = 0.007$) in rural women. Thiamine, zinc and plant protein driven nutrient pattern was associated with significant reductions in glycated hemoglobin and fasting glucose of ($\beta = -0.288$ (-0.543; -0.033) $p = 0.027$) and ($\beta = -0.382$ (-0.752; -0.012); $p = 0.043$) in rural men. Our results indicate that plant driven nutrient patterns are associated with low fasting glucose and glycated haemoglobin levels.

Keywords: plant based, fasting glucose and glycated hemoglobin, T2D, dietary patterns

1. Introduction

Type 2 diabetes (T2D) prevalence and the burden it places on populations is increasing globally thereby making it a public health challenge which requires urgent attention [1]. For instance the global prevalence of T2D among women increased from 5% to 7.9% from 1980 to 2014 [2]. The diabetes prevalence in Africa is projected to have the largest increase of 109% compared to other regions in the world by 2035 [3]. Urban black South Africans have not been spared from the T2D burden. The highest prevalence of T2D in Sub Saharan Africa was reported among urban black South Africans, with 60% of these cases being reported amongst women [4]. Black South African women have also the highest prevalence of obesity, which has been reported to be rising together with T2D in this population [5].

The adoption of Westernised lifestyles by urban dwellers is suggested to be among the leading factors resulting in the increase in non-communicable diseases NCDs such as T2D in developing countries [6]. The migration of populations from rural to urban areas has been accompanied by an increase in meat consumption as well as an increase in the consumption of sugary foods in South Africa [7]. Both of these foods are recognised as dietary risk factors for T2D [6]. However, improvements in micronutrient intake among black South African women in urban areas is suggested to be as a result of the increased consumption of fruits and vegetables which are protective of T2D risk [8-10]. Thus, the role of diet as a risk factor for T2D is complex among the black South African women. Added to this, people eat meals with a variety of nutrients which have interactive and synergistic effects on health [11]. Therefore it is difficult to determine the separate effect of a food or nutrient on disease development as it is highly interrelated with other nutrients [11]. There is need of dietary pattern analysis methods which are able to evaluate the diet as a whole and clarify the effects of the consumption of sugary foods and meat products, together with improved intakes of fruits and vegetables, relates to T2D risk amongst this population group.

Dietary pattern (or food pattern) approaches comprise of data driven methods such as factor analysis, which allow the dietary information (or food intake) at hand to determine the unique dietary pattern for the population group being evaluated [10]. Although, dietary patterns have been associated with disease risk, their effect is considered to be through nutrient intake, therefore it is pivotal to determine nutrient patterns which are associated with T2D risk [12]. This information will aid in the understanding of the aetiology of T2D. The foods that people eat are governed by their cultural norms and beliefs which vary amongst ethnic groups and which makes dietary patterns limited and not applicable across divergent population groups [12]. Evidence exists that the Western dietary pattern association with fasting glucose varies among different ethnic groups [13]. However, nutrients are universal, thereby making nutrient patterns associations with disease risk applicable for multiple population groups [14]. To the best of our knowledge, no study has reported on the association of the nutrient patterns with fasting glucose and glycated hemoglobin levels among apparently healthy individuals. T2D is a complex and multifactorial diseases, however glycated hemoglobin and fasting glucose levels are proxies for the development of this disease which are affected by diet [15]. Fasting glucose levels are indicative of the short term changes in glucose metabolism, while glycated hemoglobin depicts the long term changes [15]. This study seeks to evaluate the association of nutrient patterns derived by factor analysis with fasting glucose and glycated hemoglobin levels among apparently healthy black South Africans.

2. Materials and Methods

Study population

The study participants (n = 2010) were recruited from two urban and rural areas of the North West province into the South African arm of the Prospective urban and rural epidemiological (PURE) study using a population based sampling strategy. Apparently healthy male and female volunteers between the ages of 35 and 60 years were recruited by the fieldworkers. Individuals were considered to be apparently healthy if they were not using any medication for chronic disease and if they were not diagnosed with a chronic medical condition/disease. The international PURE study is a large-scale epidemiological study, which comprises of research

participants recruited from 17 low, middle, and high income countries [16]. The South African arm of the PURE study was initiated in 2005 with initial five-year follow-up intervals up to 2015. At baseline in 2005 and at the five year intervals during the course of the study, medical history, lifestyle behaviour (physical activity and dietary intake), blood collection (for both genetic and biochemical analyses), an electrocardiogram, and anthropometric assessments were performed to determine the role of risk factors in the development of cardiovascular diseases [16]. Our study was nested in the 2005 PURE study baseline data. Stratified nutrient pattern analysis was conducted according to gender and urban/rural status among the 2010 participants of the PURE study. However, significant associations of the nutrient patterns with fasting glucose and glycated hemoglobin were noted only among the rural women and rural men. Therefore, the detailed nutrient pattern and association results of the rural participants are elaborated in the main text while results of the urban men and women are illustrated in the supplementary section.

Ethical approval

The participants gave written informed consent before participating in the study and the study was conducted according to the Declaration of Helsinki principles[17]. Ethical approval was granted by the Ethics Committee of the North-West University, Potchefstroom Campus, with ethics number NWU-00016-10-A1.

Dietary, anthropometric and physical activity assessments

The trained fieldworkers captured the dietary intake of the participants using a standardised quantitative food frequency questionnaire (QFFQ) which had been validated for the same ethnic population group of the study participants [18,19]. The reproducibility of the QFFQ was also assessed and found to be good in this specific study population [20]. The dietary intake data was coded, analysed and nutrient intakes were computed using the South African Food Composition Database [21]. Body weight 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. The heights of the subjects were determined by the PURE study research team members using a

stadiometer (IP 1465, Invicta, and London, UK). The BMI of the participants was computed using the formula: $BMI = \text{weight (kilograms)} / \text{height (metres)}^2$. The Baecke physical activity questionnaire (BPAQ) which was validated for South Africa was used to collect the physical activity information of the participants [22]. The questionnaire was used to compute a physical activity index score as described elsewhere [22,23]. This physical activity index scores were used in the multivariate regression analysis to adjust for physical activity.

Biochemical measurements

The research participants were required to fast (at least 8 hours with no food or beverages, including water before measurements) and their blood glucose levels were measured by the PURE research team. These fasting glucose levels were measured using the SYNCHRON® System from fluoride plasma. The Bio-Rad D-10™, HbA1c kit (Bio-Rad Laboratories, Inc., Hercules, France), which operates via cation exchange high performance liquid chromatography was used to assess HbA1c levels from whole blood EDTA samples. The coefficient of variation of the glycated hemoglobin and fasting glucose tests were 1.16% and 2.1% respectively.

Statistical Analysis

The statistical analysis was performed using the statistical package for social scientists (SPSS) version 23. Normality tests for the continuous variables were performed using the Q-Q plots. Twenty-five nutrients were used to determine the nutrient patterns as has been reported previously by Pisa et. al.[24]. Among these 25 nutrients, total protein was split into animal protein and plant protein; total carbohydrates were divided into total sugar, starch, total dietary fibre; total fat was categorised into saturated fat, monounsaturated fat and polyunsaturated fat. The total dietary fibre comprised of soluble and insoluble dietary fibre. Alcohol was not regarded as a nutrient and not included in deriving the nutrient pattern analysis. However, in view of the reported association of alcohol intake with glycated hemoglobin, it was adjusted for in the multivariate linear models for the association of the derived nutrient patterns with glycated hemoglobin and fasting glucose [25].

The nutrient intake variables from the quantitative food frequency questionnaire (QFFQ) were log transformed to remove bias due to variance as a result of the different measures of scale used to quantify the nutrients. These nutrients were adjusted for log alcohol free energy using the multivariate (standard) method [11]. The multivariate method was selected as it yielded more cumulative variance and interpretable nutrient patterns compared to the standard nutrient density method of adjusting for total energy intake [11]. Principal component analysis (PCA) was the factor reduction tool used to determine the nutrient patterns from the 25 selected nutrients. The PCA was performed with the variance based on the covariance matrix and Varimax rotation. The retained principal components (PC) were used to identify the nutrient patterns. The scree plot (Figure 1) was used to determine the number of PC's to retain. The nutrient patterns were named by using the nutrients with loadings greater than ± 0.47 on the PC's. The total variance explained by the retained PC's were also evaluated to determine the relevance of the extracted PC's. The PCA was a suitable data reduction approach for the nutrient data in this study as was indicated by a Kaiser-Meyer-Olkin measure of sampling adequacy of 0.911, and a Bartlett's test of sphericity which was significant at $p < 0.001$. The PC's were categorised into tertile's and ANOVA (for continuous variables) and Chi-square (for categorical variables) tests were used to determine the descriptive characteristics of the study participants across the tertile's of the extracted PC's.

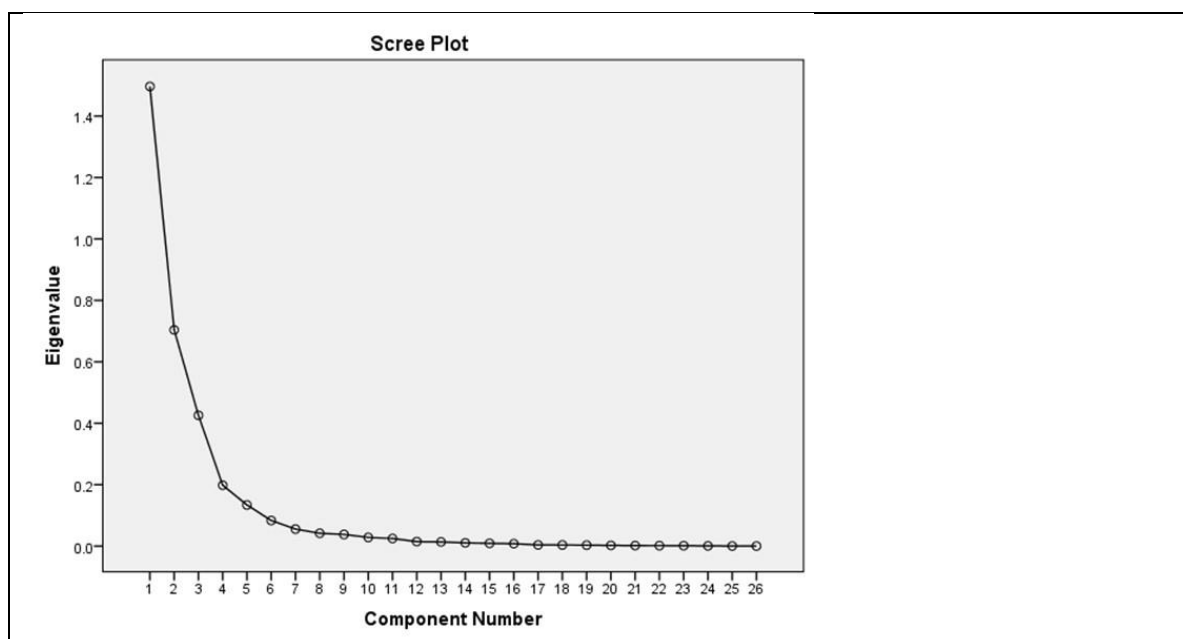


Figure 1. Scree plot of the nutrients and the extracted principal components among rural women

Crude and adjusted multiple linear regression models were computed to assess the association between the extracted PC's with glycated hemoglobin and fasting glucose as dependent variables separately through varied models. In these models, regression coefficients for 1SD increase in the PC scores (and their 95% confidence intervals) were computed for four models: M1: (crude), M2: (adjusted for M1 plus Log Total Energy), M3 : (Adjusted for M2 plus Body Mass Index), M4: (Adjusted for M3 plus age, smoking, physical activity and education level, seasonality and alcohol intake and other PC's). Seasonality was adjusted based on the months in which the dietary assessments were conducted. Thus for the rural participants seasonality was adjusted basing on August and September; and the months of October and November were used to adjust for seasonality in the urban participants. Partial R² values were computed to express the variance explained by each model. Statistical significance was regarded at p- value less than 0.05.

3. Results

3.1. Nutrient Patterns

Three nutrient patterns (Table S1) were extracted from the PCA, which explained 73% of the total variation of the selected nutrient factors among the rural women. The first nutrient pattern was depicted as “Magnesium, phosphorus and plant protein driven nutrients”. This pattern consisted of higher loadings magnesium, phosphorus and plant protein. The second nutrient pattern was termed “Fat and animal protein driven nutrients” as it had higher loadings of cholesterol, monounsaturated fat, animal protein, polyunsaturated fat and saturated fat. The “Starch, dietary fibre and B vitamins driven” based nutrient pattern was the third extracted nutrient pattern. This nutrient pattern had high loadings of starch, folate, vitamin B6, dietary fibre and thiamine as illustrated in Table S1.

Three nutrient patterns were extracted among rural men and named according to the nutrients with the highest loadings as indicated in Table S1. These were the “Thiamine, zinc and plant protein driven nutrients”, “Fat and animal protein driven nutrients”, “Retinol and vitamin B12 driven nutrients”. Three similar nutrients

patterns were extracted among the rural men, urban men and urban women which explained 76%, 77% and 76% (Table S1 and Table S2) variance respectively of the nutrients. However, some considerable differences were noted in the plant driven nutrient patterns which accounted for the greater variation of the nutrients among the urban and rural participants which are indicated in Table S3.

3.2. Descriptive characteristics of the study population

The study participants consisted of 659 rural women, 347 rural men, 605 urban women and 399 urban men. The mean fasting glucose levels and glycated hemoglobin of 4.96 ± 1.57 and 5.64 ± 0.88 respectively [26]. The women had a mean BMI of 26.73 ± 7.21 which signified that a considerable proportion of the women were overweight. Overall, in the whole study population 39.4% of the participants were overweight and 87.3% of these were women. The mean total energy intake was 7707.13 ± 3692.36 KJ. The high scores of the magnesium, phosphorus and plant protein driven nutrient pattern were associated with high total energy intake and energy from alcohol in rural women as indicated in Table 1. The high scores for the fat and animal protein driven nutrient pattern was significantly associated with tertiary education status as indicated in Table 1

Table 1. Descriptive characteristics for the rural women of the study population according to the lowest and highest tertile's of the three nutrient patterns (PC)

	Magnesium, phosphorus and plant protein driven nutrients			Fat and animal protein driven nutrients			Starch, dietary fibre and B vitamin driven nutrients		
	T1	T3	P	T1	T3	P	T1	T3	P
Age	48.72±10.23	47.51±9.57	0.247	47.94±8.89	47.73±10.16	0.837	49.15±10.40	46.94±8.99	0.036
Body Mass Index	25.28±6.63	25.27±6.58	0.988	25.31±6.75	26.13±6.85	0.262	25.26±6.41	25.65±7.11	0.605
Total energy	4322.96	8365.01	<0.001	5631.33	6883.88	<0.001	5893.08	6567.51	0.013
	±1217.30	±2476.92		±2816.62	±2189.38		±2645.21	±2425.60	
Alcohol (%TE)	0.76±2.56	7.42±12.56	<0.001	5.54±11.70	1.47±4.40	<0.001	5.36±11.91	1.69±5.01	<0.001
Protein (%TE)	11.26±1.96	10.77±1.64	0.008	10.27±1.33	11.67±1.81	<0.001	11.23±2.05	10.91±1.31	0.092
Current Smokers (%)	33.5	36.8	0.168	38.0	34.7	0.005	30.6	36.4	0.595
Physical Activity Index	8.19±1.42	8.28±1.42	0.593	8.29±1.29	8.32±1.55	0.854	8.23±1.37	8.44±1.45	0.205
Tertiary education (%)	33.7	24.5	0.058	26.5	42.9	0.028	28.6	38.8	0.854
Fasting glucose (mmolL⁻¹)	4.73 ± 0.78	4.93 ± 1.47	0.219	4.91 ± 1.18	4.87 ± 1.02	0.784	5.12 ± 2.41	4.88 ± 0.83	0.160
HbA1C (%)	5.64 ± 0.53	5.67 ± 0.93	0.785	5.68 ± 0.91	5.69 ± 0.58	0.843	5.84 ± 1.23	5.62 ± 0.55	0.027

P= p value based on ANOVA or Chi- square test where appropriate; Alcohol (%TE) =percentage of total energy due to alcohol intake; Protein (%TE) =percentage of total energy due to protein intake; T1= lowest tertile; T3= highest tertile

High scores of the thiamine, zinc and plant protein nutrient pattern were associated with higher energy and alcohol intake compared to other nutrient patterns among the rural men in Table 2.

Table 2. Descriptive characteristics for the rural men of the study population according to the lowest and highest tertile's of the three nutrient patterns (PC)

	Thiamine, zinc and plant protein driven nutrients			Fat and animal protein driven nutrients			Retinol and vitamin B12 driven nutrients		
	T1	T3	P	T1	T3	P	T1	T3	P
Age	47.95±9.99	50.19±10.69	0.156	48.43±9.97	51.42±11.45	0.057	49.66±10.01	49.57±9.95	0.954
Body Mass Index	20.86±4.15	20.54±4.31	0.606	20.86±3.99	20.94±4.65	0.894	20.30±3.57	20.95±4.19	0.297
Total energy	4693.60	10637.00	<0.001	6319.63	8220.85	<0.001	7164.14	7855.83	0.159
	±1584.26	±2887.76		±3228.51	±3228.51		±2975.31	±3672.65	
Alcohol (%TE)	5.98±9.40	13.82±13.70	<0.001	11.19±13.41	5.70±9.16	0.002	6.65±11.23	9.19±11.79	0.159
Protein (%TE)	11.51±2.69	10.76±1.53	0.014	10.09±1.43	12.19±2.33	<0.001	10.64±1.69	11.50±2.48	0.005
Current Smokers (%)	32.4	35.9	0.636	40.0	26.2	0.003	31.7	31.7	0.577
Physical Activity Index	8.25±1.64	8.02±1.44	0.380	8.03±1.65	7.99±1.79	0.861	7.92±1.61	7.97±1.63	0.847
Tertiary education (%)	44.2	20.9	0.141	23.3	37.2	0.514	27.9	37.2	0.451
Fasting glucose (mmolL⁻¹)	4.90 ± 0.93	4.68 ± 0.90	0.113	4.75 ± 0.79	4.96 ± 1.24	0.138	4.94 ± 2.41	4.75 ± 0.78	0.190
HbA1C (%)	5.59 ± 0.52	5.51 ± 0.81	0.386	5.49 ± 0.34	5.62 ± 0.97	0.169	5.57 ± 0.83	5.53 ± 0.49	0.678

, P= p value based on ANOVA or Chi- square test where appropriate; Alcohol (%TE) =percentage of total energy due to alcohol intake; Protein (%TE) =percentage of total energy due to protein intake; T1= lowest tertile; T3= highest tertile

3.3. Nutrient patterns associations with fasting and glycated hemoglobin levels

The association results of glycated hemoglobin and fasting glucose levels with 1 SD increases in the extracted nutrient patterns are shown in Table 3 to Table 6 for the rural women and rural men. The magnesium, phosphorus and plant protein driven nutrient pattern

was associated with a consistent trend of increases in fasting glucose and glycated hemoglobin among rural women as illustrated in Table 3.

Table 3. Regression coefficients for fasting glucose for 1 SD increase in the derived nutrient pattern scores among rural black South African women

	Magnesium, phosphorus and plant protein driven nutrients			Fat and animal protein driven nutrients			Starch, dietary fibre and b vitamin driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	0.129 (-0.014; 0.271)	0,077	0.007	0.009 (-0.141; 0.160)	0,902	0.000	-0.164 (-0.311; -0.018)	0,027	0.008
M2	0.196 (-0.063; 0.455)	0,138	0.007	-0.020 (-0.183; 0.143)	0,813	0.003	-0.197 (-0.349; -0.049)	0,011	0.016
M3	0.278 (-0.001; 0.280)	0,034	0.045	-0.038 (-0.198; 0.123)	0,645	0.037	-0.203 (-0.351; -0.054)	0,008	0.051
M4	0.147 (-0.360; 0.655)	0.569	0.086	-0.004 (-0.290; 0.281)	0,976	0.086	-0.236 (-0.458; -0.014)	0,037	0.086

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 : (Adjusted for M2 plus Body Mass Index); M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients Fasting glucose units mmolL⁻¹ = millimoles per litre.

Conversely. the starch, dietary fibre and B vitamins nutrient pattern was constantly associated with reduced glycated hemoglobin and fasting glucose levels of in all the evaluated linear models, for instance the M4 model: -0.175 % (-0.303; -0.047); p =0.007) and -0.236 mmolL⁻¹ (-0.458; -0.014); p = 0.037) respectively as indicated in Table 3 and Table 4 among rural women.

Table 4. Regression coefficients for glycated hemoglobin for 1 SD increase in the derived nutrient pattern scores among rural black South African women

	Magnesium, phosphorus and plant protein driven nutrients			Fat and animal protein driven nutrients			Starch, dietary fibre and b-vitamin driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	0.029 (-0.055; 0.112)	0.502	0.001	0.032 (-0.056; 0.120)	0.477	0.001	-0.138 (-0.224; -0.053)	0.002	0.020
M2	0.048 (-0.104; 0.199)	0.538	0.001	0.028 (-0.067; 0.123)	0.563	0.001	-0.145 (-0.234; -0.056)	0.001	0.021
M3	0.112 (-0.036; 0.260)	0.139	0.069	0.014 (-0.078; 0.106)	0.766	0.065	-0.151 (-0.237; -0.065)	0.001	0.087
M4	0.107 (-0.188; 0.401)	0.478	0.150	-0.011 (-0.175; 0.154)	0.478	0.150	-0.175 (-0.303; -0.047)	0.007	0.150

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 :(Adjusted for M2 plus Body Mass Index); M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients. Glycated hemoglobin unit=%

The thiamine, zinc and plant protein nutrient patterns was associated with reduced glycated hemoglobin and fasting glucose levels in rural men as illustrated in Table 5 and Table 6. For instance, the M4 models which explained the highest variances of the fasting glucose and glycated hemoglobin indicated reductions of these outcome variables to be -0.382 mmolL⁻¹ (-0.752; -0.012; p =0.043) and -0.288% (-0.543; -0.033; p =0.027) respectively as illustrated in Table 5 and 6.

Table 5. Regression coefficients for fasting glucose for 1 SD increase in the derived nutrient pattern scores among rural black South African men

	Thiamine, zinc and plant protein driven nutrients			Fat and animal protein driven nutrients			Retinol and vitamin B12 driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	-0.057 (-0.172; 0.057)	0.326	0.004	0.054 (-0.061; 0.169)	0,355	0.003	-0.039 (-0.156; 0.077)	0,504	0.002
M2	-0.237 (-0.492; 0.019)	0.069	0.013	0.061 (-0.064; 0.186)	0,335	0.004	-0.055 (-0.173; 0.064)	0,363	0.003
M3	-0.255 (-0.496; 0.014)	0.038	0.117	0.055 (-0.063; 0.172)	0,363	0.115	-0.082 (-0.194; 0.030)	0,153	0.120
M4	-0.382 (-0.752; -0.012)	0.043	0.182	-0.051 (-0.241; 0.139)	0,596	0.182	-0.109 (-0.229; 0.032)	0,074	0.182

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 :(Adjusted for M2 plus Body Mass Index): M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients Fasting glucose units mmolL⁻¹ = millimoles per litre.

The plant driven nutrient patterns had significant results as has been shown above among the urban and rural participants. However, some notable differences were depicted in these plant driven nutrient patterns illustrated in Table S3

Table 6. Regression coefficients for glycated hemoglobin for 1 SD increase in the derived nutrient pattern scores among rural black South African women

	Thiamine, zinc and plant protein driven nutrients			Fat and animal protein driven nutrients			Retinol and vitamin b12 driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	-0.039 (-0.117; 0.038)	0,320	0.000	0.046 (-0.031; 0.123)	0,241	0.005	0.012 (-0.065; 0.090)	0.754	0.000
M2	-0.214 (-0.384; 0.044)	0,014	0.024	0.053 (-0.031; 0.138)	0,213	0.006	0.015 (-0.065; 0.094)	0,713	0.001
M3	-0.230 (-0.392; -0.067)	0,006	0.113	0.050 (-0.030; 0.131)	0,219	0.092	0.001 (-0.075; 0.077)	0,975	0.086
M4	-0.288 (-0.543; -0.033)	0.027	0.174	-0.057 (-0.189; 0.075)	0,396	0.174	-0.018 (-0.100; 0.064)	0,662	0.174

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 :(Adjusted for M2 plus Body Mass Index): M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients. Glycated hemoglobin unit=%

The following nutrient patterns, “thiamine, zinc and plant protein driven nutrients”; “thiamine, starch and folate driven nutrients” and “magnesium, phosphorus and plant protein driven nutrients” extracted from the urban men, urban women and rural women respectively, had higher loadings for animal protein, saturated fat, mono-saturated fat and sugar compared to the “thiamine, zinc and plant protein driven nutrients” and “the starch, dietary fibre and B vitamin driven nutrients patterns” which were associated with significant reductions in fasting and glycated hemoglobin among the rural men and women respectively (see factor loadings in bold in Table S3). These two nutrient patterns that had significant associations with fasting glucose and glycated hemoglobin had very low cholesterol and saturated fat as illustrated in Table S3.

4. Discussion

We set out to determine the nutrient patterns associated with fasting glucose and glycated hemoglobin levels among apparently healthy volunteers. The principal component analysis method enabled the extraction of three nutrient patterns among a black South African population which explained about 73% of the variation of the nutrient factors in the urban/rural and gender stratifications. The magnesium, phosphorus and plant protein driven nutrient pattern was associated with a trend of increasing fasting glucose and glycated hemoglobin levels per 1 SD increase in the pattern in the rural women while the thiamine, zinc and plant protein nutrient pattern was associated with a positive trend of increasing glycated hemoglobin among urban men. Notably, the starch, dietary fibre and B vitamin nutrient pattern was associated with decreases in glycated hemoglobin and fasting glucose levels - 0.175 % (-0.303; -0.047); $p=0.007$) and $-0.236 \text{ mmolL}^{-1}$ (-0.458; -0.014); $p=0.037$) respectively among rural women. The thiamine, zinc and plant protein driven nutrient pattern was associated with significant reductions in fasting glucose and glycated hemoglobin of $-0.382 \text{ mmolL}^{-1}$ (-0.752; -0.012; $p=0.043$) and -0.288% (-0.543; -0.033; $p=0.027$) in rural men. These associations were significantly maintained after adjusting for age, BMI, log total energy intake, smoking, physical activity, alcohol intake, seasonality and education level attained thus indicating an independent association of the starch, dietary fibre and B vitamin nutrient pattern and the thiamine, zinc and plant protein driven nutrient pattern with fasting glucose and glycated hemoglobin in the rural women and rural men strata's respectively.

Comparable studies evaluating the associations of nutrients patterns with fasting glucose and glycated hemoglobin are scarce. However, evidence exists of dietary patterns which have evaluated this phenomenon [10]. The Health/Prudent dietary pattern has been associated with decreases in fasting glucose and glycated hemoglobin levels while the Western dietary pattern has been associated with increases in these biomarkers of T2D [27,28]. However, the association of the Western dietary pattern which is characterised with high intakes of animal proteins and snacks with fasting glucose levels was noted to vary among ethnic groups [13]. The Western dietary pattern was reported in one study to be only significantly

associated with high fasting glucose levels and glycated hemoglobin levels among Dutch and not among the Moroccans and Turkish groups [13]. Dietary patterns are known to vary among different ethnic groups, therefore nutrient patterns that were considered in our study are regarded to aid in indicating a non-ethnic insight into this phenomenon.

In our study we noted that fat and animal protein driven nutrients were also not associated with fasting glucose and glycated levels as had been depicted in dietary pattern analysis studies among different ethnic groups [13]. This disparity from the Western populations where the animal based/Western dietary patterns are associated with increasing fasting glucose levels has been explained by the realisation that ethnic groups such as Asians may adopt the Western diets but their intake of animal products will remain lower and as they also continue to consume traditional cereals and vegetables [29]. Similarly, in this study population, the intake of fat and animal protein nutrients were lower in the plant driven nutrients patterns that accounted for the greatest variance among the three nutrients patterns extracted per stratum as illustrated in Table S3. In other local studies, the protein and fat intakes as percentage of total energy intake for urban women from 1975 to 2005 did not change drastically though evidences of the adoption of Western dietary patterns were being noted and the greater proportion of the energy intake was still being contributed by the carbohydrate intake as had been noted in the Asians [6]. From 1975 to 2005 the percentage of energy of protein intake changed from 14% to 13%, while fat intake as percentage of energy changed from 21% to 30% and carbohydrate intake changed from 67% and 57% of total energy intake [6]. Thus, the intake of fat and animal protein driven nutrients in this population group might have been lower and thus not associated with increases in fasting glucose and glycated hemoglobin as was expected.

The magnesium, phosphorus and plant protein driven nutrient pattern indicated a positive trend association with increases in fasting glucose and glycated hemoglobin levels among rural women, while the thiamine, zinc and plant protein driven nutrient pattern was associated with a trend of increasing glycated hemoglobin levels in the urban men (Table S7). However, plant protein and zinc

which were high in these nutrient pattern have been reported to independently lower fasting glucose levels [30-33]. In addition, the thiamine, zinc and plant protein driven nutrient pattern was associated with a trend of decreasing glycated hemoglobin and fasting glucose levels in the rural men. In view, that the study population comprised of people who consumed diets of both animal and plant based nutrients and not purely vegetarians, the discrepancies in the associations of the plant driven nutrients can be explained by the varied proportions of animal protein, saturated fat, mono-saturated fat, cholesterol and sugar among these plant driven nutrient patterns as illustrated in Table S3. The magnesium, phosphorus and plant protein driven nutrient pattern in rural women and the thiamine, zinc and plant protein driven nutrient pattern in urban men had higher loadings of animal protein, saturated fat, mono-saturated fat, cholesterol and sugar compared to the other nutrient patterns discussed below such as the starch, dietary fibre and B vitamins driven nutrient pattern among rural women and the thiamine, zinc and plant protein nutrient pattern among rural men which associated with low fasting glucose and glycated hemoglobin levels. Animal protein, saturated fat, cholesterol and sugar are high in Western dietary patterns which have been previously associated with increased risk of T2D and this helps clarify the positive trend of association of the magnesium, phosphorus and plant protein driven nutrient pattern in rural women and the thiamine, zinc and plant protein driven nutrient pattern in urban men with the study outcome variables [34]. The comparisons of the varied constituents of the plant driven nutrients as discussed above and illustrated in Table S3 indicates the attractiveness of the PCA approach of nutrient pattern determination as it allows a whole based approach of the effect of nutrients to particular outcomes to be evaluated.

The starch, dietary fibre and B vitamins driven nutrient pattern was consistently associated with reduced fasting glucose and glycated hemoglobin levels in all the multivariate linear models which were considered in this study among rural women. The thiamine, zinc and plant protein nutrient pattern was also associated with significant reductions in glycated hemoglobin and fasting glucose among rural men. These findings are similar to the results of other studies on plant based dietary patterns and T2D risk [10,34]. It has been consistently reported that plant based

diets are protective against T2D risk [10,34,35]. A number of mechanisms have been proposed to explain this phenomenon [10]. Plant based diets which are rich in dietary fibre are suggested to reduce T2D risk by reducing postprandial insulin demand, improving insulin sensitivity and the antioxidants found in these diets may help enhance β -cell function [36,37]. Dietary fibre from cereal foods which are also high in starch has been consistently shown across eight European countries to be associated with reduced T2D risk [38]. Maize meal fortified with iron, zinc and B vitamins which is largely consumed among the black South African population group is known to contain resistant starches that are partially digested and have been associated with improved insulin sensitivity which may then lead to reductions in fasting glucose levels [39,40]. Thiamine was also high in this nutrient pattern. Evidence exists that dietary fibre beneficial effects might also be due to its concomitant intake together with thiamine [41]. The results of this study suggest that starch, dietary fibre and B vitamins nutrients, zinc and plant protein consumed together may lead to the lowering of fasting and glycated hemoglobin levels as has been postulated elsewhere [35].

The strengths of the current study include the use of a validated QFFQ and recruitment criteria of selecting apparently healthy individuals who were not taking chronic medications or suffering from chronic diseases. This might have helped to prevent the confounding of drugs and T2D to the nutrient pattern association results with glycated hemoglobin and fasting glucose levels. The factor analysis approach used to derive the nutrient patterns is known for depicting real-world dietary behaviours [42]. However, this approach is based on a number of subjective decisions such as naming nutrient patterns, method of rotation and selection of food groups which can lead to an overall measurement error [10,42]. Since this approach is a posteriori analysis tool, it derives nutrient patterns based on data at hand and this makes comparisons with other studies difficult [10,42]. However, regardless of the differences in the constituents of the nutrient patterns due to the data driven approaches used to derive them, consistent similarities of the dietary factors associated with T2D risk as reported in this study, has been depicted in multiple populations [10,34,37]. It might be probable that the residual confounding effect of total energy intake might have led to distortions in the association of the nutrient

patterns with fasting glucose and glycated hemoglobin levels. Residual confounding is a result of measurement error in a confounder included in the model [43]. Although the nutrient and total energy intake based on the QFFQ's are all not precisely measured, adjustment for total energy intake should control for total energy intake confounding [44]. However, evidence exists that the control of total energy intake is not complete in epidemiological studies involving QFFQs thereby leading to residual confounding [44,45]. Therefore, there is need to further explore the association of the nutrient patterns in isocaloric clinical trials which are better designed to control for confounding for total energy intake [45].

5. Conclusions

In summary, our results indicate the beneficial associations of plant driven nutrient patterns with reductions in fasting glucose and glycated hemoglobin levels. However, the small variances that were explained by the models explored in this study are suggestive of the presence of other factors that affect the variation of fasting glucose and glycated hemoglobin which were not accounted for in this study. Thus more studies are required to further explore the association of nutrient patterns with fasting glucose and glycated hemoglobin.

Acknowledgments: This research was supported by a grant National Research Foundation of South Africa (87970). Tinashe Chikowore is supported by the National Research Foundation of South Africa (89117). We would like to acknowledge Prof A Kruger and the PURE research team, funders and participants. All the authors have no conflicts of interest to declare.

Author Contributions: TC conceptualized the paper, analysed the data and wrote the draft article. PTP designed the data analysis plan and contributed to the writing of the paper. EWV designed the dietary data collection tools, coded and analysed the data for nutrients and contributed to the writing of the paper. TZ, EJMF and KRC supervised and contributed to the writing of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, L.; Magliano, D.J.; Zimmet, P.Z. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. *Nat Rev Endocrinol* **2012**, *8*, 228-236.
2. Collaboration, N.C.D.R.F. Worldwide trends in diabetes since 1980: A pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **2016**, *387*, 1513-1530.
3. Guariguata, L.; Whiting, D.R.; Hambleton, I.; Beagley, J.; Linnenkamp, U.; Shaw, J.E. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes research and clinical practice* **2014**, *103*, 137-149.
4. Peer, N.; Steyn, K.; Lombard, C.; Lambert, E.V.; Vythilingum, B.; Levitt, N.S. Rising diabetes prevalence among urban-dwelling black south africans. *PloS one* **2012**, *7*, e43336.
5. Sartorius, B.; Veerman, L.J.; Manyema, M.; Chola, L.; Hofman, K. Determinants of obesity and associated population attributability, south africa: Empirical evidence from a national panel survey, 2008-2012. *PLoS One* **2015**, *10*, e0130218.
6. Vorster, H.H.; Kruger, A.; Margetts, B.M. The nutrition transition in africa: Can it be steered into a more positive direction? *Nutrients* **2011**, *3*, 429-441.
7. Vorster, H.H.; Kruger, A.; Wentzel-Viljoen, E.; Kruger, H.S.; Margetts, B.M. Added sugar intake in south africa: Findings from the adult prospective urban and rural epidemiology cohort study. *Am J Clin Nutr* **2014**, *99*, 1479-1486.
8. Vorster, H.H.; Margetts, B.M.; Venter, C.S.; Wissing, M.P. Integrated nutrition science: From theory to practice in south africa. *Public Health Nutr* **2005**, *8*, 760-765.
9. Hattingh, Z.; Walsh, C.M.; Bester, C.J.; Oguntibeju, O.O. Evaluation of energy and macronutrient intake of black women in bloemfontein: A cross-sectional study. *Afr J Biotechnol* **2008**, *7*, 4019-4024.
10. McEvoy, C.T.; Cardwell, C.R.; Woodside, J.V.; Young, I.S.; Hunter, S.J.; McKinley, M.C. A posteriori dietary patterns are related to risk of type 2 diabetes: Findings from a systematic review and meta-analysis. *J Acad Nutr Diet* **2014**, *114*, 1759-1775 e1754.
11. Hu, F.B. Dietary pattern analysis: A new direction in nutritional epidemiology. *Curr Opin Lipidol* **2002**, *13*, 3-9.
12. Salehi-Abargouei, A.; Esmailzadeh, A.; Azadbakht, L.; Keshteli, A.H.; Feizi, A.; Feinle-Bisset, C.; Adibi, P. Nutrient patterns and their relation to general and abdominal obesity in iranian adults: Findings from the sepanhan study. *Eur J Nutr* **2016**, *55*, 505-518.

13. Dekker, L.H.; van Dam, R.M.; Snijder, M.B.; Peters, R.J.; Dekker, J.M.; de Vries, J.H.; de Boer, E.J.; Schulze, M.B.; Stronks, K.; Nicolaou, M. Comparable dietary patterns describe dietary behavior across ethnic groups in the netherlands, but different elements in the diet are associated with glycated hemoglobin and fasting glucose concentrations. *J Nutr* **2015**, *145*, 1884-1891.
14. Freisling, H.; Fahey, M.T.; Moskal, A.; Ocke, M.C.; Ferrari, P.; Jenab, M.; Norat, T.; Naska, A.; Welch, A.A.; Navarro, C., *et al.* Region-specific nutrient intake patterns exhibit a geographical gradient within and between european countries. *J Nutr* **2010**, *140*, 1280-1286.
15. Koenig, R.J.; Peterson, C.M.; Jones, R.L.; Saudek, C.; Lehrman, M.; Cerami, A. Correlation of glucose regulation and hemoglobin a(1c) in diabetes mellitus. *New England Journal of Medicine* **1976**, *295*, 417-420.
16. Teo, K.; Chow, C.K.; Vaz, M.; Rangarajan, S.; Yusuf, S.; Group, P.I.-W. The prospective urban rural epidemiology (pure) study: Examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. *American heart journal* **2009**, *158*, 1-7 e1.
17. Arie, S. Revision of helsinki declaration aims to prevent exploitation of study participants. *Bmj-Brit Med J* **2013**, *347*.
18. MacIntyre, U.E.; Venter, C.S.; Vorster, H.H.; Steyn, H.S. 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* **2001**, *4*, 45-51.
19. Wolmarans, P.; Danster, N.; Dalton, A.; Rossouw, K.; Schönfeldt, H. Condensed food composition tables for south africa. *Cape Town: Medical Research Council* **2010**.
20. Wentzel-Viljoen, E.; Laubscher, R.; Kruger, A. Using different approaches to assess the reproducibility of a culturally sensitive quantified food frequency questionnaire. *South African journal of clinical nutrition* **2011**, *24*, 143-148.
21. Wolmarans, P.; Chetty, J.; Danster-Christians, N. Food composition activities in south africa. *Food Chem* **2013**, *140*, 447-450.
22. Kruger, H.; Venter, C.; Steyn, H. A standardised physical activity questionnaire for a population in transition: The thusa study. *Afr J Phys Health Educ Recreat Dance* **2000**, *6*, 54-64.
23. Oyeyemi, A.L.; Moss, S.J.; Monyeki, M.A.; Kruger, H.S. Measurement of physical activity in urban and rural south african adults: A comparison of two self-report methods. *BMC Public Health* **2016**, *16*, 1004.
24. Pisa, P.T.; Pedro, T.M.; Kahn, K.; Tollman, S.M.; Pettifor, J.M.; Norris, S.A. Nutrient patterns and their association with socio-demographic, lifestyle factors and obesity risk in rural south african adolescents. *Nutrients* **2015**, *7*, 3464-3482.
25. Boeing, H.; Weisgerber, U.M.; Jeckel, A.; Rose, H.J.; Kroke, A. Association between glycated hemoglobin and diet and other lifestyle factors in a nondiabetic population: Cross-sectional evaluation of data from the potsdam cohort of the

- European prospective investigation into cancer and nutrition study. *Am J Clin Nutr* **2000**, *71*, 1115-1122.
26. Alberti, K.G.; Zimmet, P.Z. 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* **1998**, *15*, 539-553.
 27. Kim, H.S.; Park, S.Y.; Grandinetti, A.; Holck, P.S.; Waslien, C. Major dietary patterns, ethnicity, and prevalence of type 2 diabetes in rural Hawaii. *Nutrition* **2008**, *24*, 1065-1072.
 28. Liese, A.D.; Weis, K.E.; Schulz, M.; Toozé, J.A. Food intake patterns associated with incident type 2 diabetes: The insulin resistance atherosclerosis study. *Diabetes care* **2009**, *32*, 263-268.
 29. Gilbert, P.A.; Khokhar, S. Changing dietary habits of ethnic groups in Europe and implications for health. *Nutr Rev* **2008**, *66*, 203-215.
 30. Roberts-Toler, C.; O'Neill, B.T.; Cypess, A.M. Diet-induced obesity causes insulin resistance in mouse brown adipose tissue. *Obesity (Silver Spring)* **2015**, *23*, 1765-1770.
 31. Capdor, J.; Foster, M.; Petocz, P.; Samman, S. Zinc and glycemic control: A meta-analysis of randomised placebo controlled supplementation trials in humans. *J Trace Elem Med Biol* **2013**, *27*, 137-142.
 32. Hruby, A.; Ngwa, J.S.; Renstrom, F.; Wojczynski, M.K.; Ganna, A.; Hallmans, G.; Houston, D.K.; Jacques, P.F.; Kanoni, S.; Lehtimäki, T., *et al.* Higher magnesium intake is associated with lower fasting glucose and insulin, with no evidence of interaction with select genetic loci, in a meta-analysis of 15 consortia studies. *J Nutr* **2013**, *143*, 345-353.
 33. Khattab, M.; Abi-Rashed, C.; Ghattas, H.; Hlais, S.; Obeid, O. Phosphorus ingestion improves oral glucose tolerance of healthy male subjects: A crossover experiment. *Nutr J* **2015**, *14*, 112.
 34. Alhazmi, A.; Stojanovski, E.; McEvoy, M.; Garg, M.L. The association between dietary patterns and type 2 diabetes: A systematic review and meta-analysis of cohort studies. *J Hum Nutr Diet* **2014**, *27*, 251-260.
 35. Sabate, J.; Wien, M. A perspective on vegetarian dietary patterns and risk of metabolic syndrome. *Br J Nutr* **2015**, *113 Suppl 2*, S136-143.
 36. Barclay, A.W.; Petocz, P.; McMillan-Price, J.; Flood, V.M.; Prvan, T.; Mitchell, P.; Brand-Miller, J.C. Glycemic index, glycemic load, and chronic disease risk--a meta-analysis of observational studies. *Am J Clin Nutr* **2008**, *87*, 627-637.
 37. Montonen, J.; Knekt, P.; Harkanen, T.; Jarvinen, R.; Heliovaara, M.; Aromaa, A.; Reunanen, A. Dietary patterns and the incidence of type 2 diabetes. *American journal of epidemiology* **2005**, *161*, 219-227.
 38. InterAct, C. Dietary fibre and incidence of type 2 diabetes in eight European countries: The epic-interact study and a meta-analysis of prospective studies. *Diabetologia* **2015**, *58*, 1394-1408.

39. Brites, C.M.; Trigo, M.J.; Carrapico, B.; Alvina, M.; Bessa, R.J. Maize and resistant starch enriched breads reduce postprandial glycemic responses in rats. *Nutr Res* **2011**, *31*, 302-308.
40. Gower, B.A.; Bergman, R.; Stefanovski, D.; Darnell, B.; Ovalle, F.; Fisher, G.; Sweatt, S.K.; Resuehr, H.S.; Pelkman, C. Baseline insulin sensitivity affects response to high-amylose maize resistant starch in women: A randomized, controlled trial. *Nutr Metab (Lond)* **2016**, *13*, 2.
41. Bakker, S.J.; Hoogeveen, E.K.; Nijpels, G.; Kostense, P.J.; Dekker, J.M.; Gans, R.O.; Heine, R.J. The association of dietary fibres with glucose tolerance is partly explained by concomitant intake of thiamine: The hoorn study. *Diabetologia* **1998**, *41*, 1168-1175.
42. Moeller, S.M.; Reedy, J.; Millen, A.E.; Dixon, L.B.; Newby, P.K.; Tucker, K.L.; Krebs-Smith, S.M.; Guenther, P.M. Dietary patterns: Challenges and opportunities in dietary patterns research an experimental biology workshop, april 1, 2006. *J Am Diet Assoc* **2007**, *107*, 1233-1239.
43. Fewell, Z.; Davey Smith, G.; Sterne, J.A. The impact of residual and unmeasured confounding in epidemiologic studies: A simulation study. *American journal of epidemiology* **2007**, *166*, 646-655.
44. Willett, W. *Nutritional epidemiology*. Oxford University Press: New York, 1998.
45. Jakes, R.W.; Day, N.E.; Luben, R.; Welch, A.; Bingham, S.; Mitchell, J.; Hennings, S.; Rennie, K.; Wareham, N.J. Adjusting for energy intake--what measure to use in nutritional epidemiological studies? *Int J Epidemiol* **2004**, *33*, 1382-1386.

SUPPLEMENT SECTION

Table S1. Extracted nutrient patterns and factor loadings of rural women and rural men

	Rural women			Rural Men		
Nutrients	Magnesium, phosphorus and plant protein driven nutrients	Fat and animal protein driven nutrients	Starch, dietary fibre and B vitamin driven nutrients	Thiamine, zinc and plant protein driven nutrients	Fat and animal protein driven nutrients	Retinol and vitamin B12 driven nutrients
Plant protein	0.875	0.211	0.301	0.924	0.165	-0.030
Animal protein	0.260	0.831	0.072	0.153	0.781	0.371
Saturated fat	0.343	0.807	-0.037	0.228	0.830	0.209
Monounsaturated fat	0.326	0.876	0.008	0.252	0.886	0.117
Polyunsaturated fat	0.301	0.855	0.099	0.323	0.830	-0.122
Cholesterol	0.137	0.891	0.148	-0.030	0.827	0.338
Starch	0.173	0.071	0.949	0.758	-0.099	-0.138
Sugar	0.270	0.352	0.042	0.164	0.274	0.218
Dietary fibre	0.680	0.208	0.608	0.891	0.157	-0.086
Calcium	0.685	0.424	0.025	0.637	0.321	0.437
Iron	0.853	0.181	0.394	0.924	0.145	-0.028
Magnesium	0.943	0.117	0.122	0.878	0.072	0.099
Phosphorus	0.886	0.334	0.138	0.850	0.267	0.204
Potassium	0.842	0.416	0.059	0.836	0.340	0.185
Zinc	0.855	0.285	0.337	0.927	0.261	0.043
Retinol	-0.012	0.261	-0.106	-0.082	0.145	0.981
Beta carotene	0.080	0.460	0.059	-0.013	0.311	0.177
Thiamine	0.775	0.043	0.546	0.952	0.028	-0.086
Riboflavin	0.798	0.305	0.227	0.792	0.157	0.325
Vitamin B6	0.520	0.181	0.756	0.829	0.108	-0.043
Folate	0.403	0.135	0.859	0.781	0.060	-0.080
Vitamin B12	0.154	0.539	0.157	-0.023	0.436	0.569
Vitamin C	0.331	0.605	0.147	0.254	0.548	0.065
Vitamin D	0.101	0.803	0.097	-0.092	0.785	0.161
Vitamin E	0.190	0.809	0.193	0.237	0.769	-0.144
Explained variance%	32.747	27.116	13.536	40.411	23.127	8.336
Cumulative explained variance %	32.747	59.863	73.399	40.411	63.539	76.180

Bold factor loadings used to indicate factor loadings $> \pm 0.47$ for naming the nutrient patterns

Table S2 Nutrient patterns and factor loadings for urban participants

	Urban Women			Urban Man		
Nutrients	Thiamine, starch and folate driven nutrients	Fat and animal protein driven nutrients	Retinol, Vitamin B12 and calcium driven nutrients	Thiamine, zinc and plant protein driven nutrients	Fat and animal protein driven nutrients	Beta carotene and Vitamin C driven nutrients
Plant protein	0.877	0.288	0.132	0.896	0.272	0.135
Animal protein	0.347	0.795	0.224	0.408	0.737	0.211
Saturated fat	0.363	0.746	0.290	0.412	0.741	0.251
Monounsaturated fat	0.366	0.776	0.185	0.408	0.766	0.282
Polyunsaturated fat	0.444	0.720	0.129	0.375	0.789	0.240
Cholesterol	0.261	0.827	0.252	0.217	0.831	0.103
Starch	0.898	0.057	0.016	0.898	0.108	0.095
Sugar	0.439	0.378	0.255	0.365	0.423	0.465
Dietary fibre	0.853	0.278	0.125	0.879	0.267	0.208
Calcium	0.405	0.434	0.512	0.435	0.449	0.142
Iron	0.850	0.381	0.169	0.903	0.313	0.182
Magnesium	0.841	0.320	0.164	0.863	0.240	0.186
Phosphorus	0.709	0.520	0.291	0.759	0.452	0.214
Potassium	0.684	0.498	0.267	0.736	0.446	0.327
Zinc	0.815	0.466	0.150	0.859	0.395	0.213
Retinol	0.077	0.270	0.950	0.053	0.272	0.072
Beta carotene	0.298	0.388	0.124	0.312	0.297	0.885
Thiamine	0.914	0.276	0.114	0.942	0.222	0.165
Riboflavin	0.612	0.520	0.371	0.744	0.409	0.197
Vitamin B6	0.833	0.366	0.160	0.861	0.301	0.270
Folate	0.889	0.257	0.151	0.889	0.298	0.122
Vitamin B12	0.256	0.558	0.611	0.352	0.549	0.141
Vitamin C	0.435	0.392	0.163	0.401	0.408	0.517
Vitamin D	0.150	0.763	0.213	0.129	0.804	0.054
Vitamin E	0.341	0.743	0.077	0.275	0.823	0.221
Explained variance%	38.984	27.351	9.426	41.885	26.247	8.370
Cumulative explained variance %	38.984	66.335	75.760	41.885	68.132	76.502

Bold factor loadings used to indicate factor loadings $\geq \pm 0.47$ for naming the nutrient patterns

Table S3. Comparison of the factors loadings in the plant driven nutrient patterns among the urban and rural participants

	Rural Men	Urban Men	Urban Females	Rural Women	
Nutrients	Thiamine, zinc and plant protein driven nutrients	Thiamine, zinc and plant protein driven nutrients	Thiamine, starch and folate driven nutrients	Magnesium, phosphorus and plant protein driven nutrients	Starch, dietary fibre and B vitamin driven nutrients
Plant protein	0.924	0.896	0.877	0.875	0.301
Animal protein	0.153	0.408	0.347	0.260	0.072
Saturated fat	0.228	0.412	0.363	0.343	-0.037
Monounsaturated fat	0.252	0.408	0.366	0.326	0.008
Polyunsaturated fat	0.323	0.375	0.444	0.301	0.099
Cholesterol	-0.030	0.217	0.261	0.137	0.148
Starch	0.758	0.898	0.898	0.173	0.949
Sugar	0.164	0.365	0.439	0.270	0.042
Dietary fibre	0.891	0.879	0.853	0.680	0.608
Calcium	0.637	0.435	0.405	0.685	0.025
Iron	0.924	0.903	0.850	0.853	0.394
Magnesium	0.878	0.863	0.841	0.943	0.122
Phosphorus	0.850	0.759	0.709	0.886	0.138
Potassium	0.836	0.736	0.684	0.842	0.059
Zinc	0.927	0.859	0.815	0.855	0.337
Retinol	-0.082	0.053	0.077	-0.012	-0.106
Beta carotene	-0.013	0.312	0.298	0.080	0.059
Thiamine	0.952	0.942	0.914	0.775	0.546
Riboflavin	0.792	0.744	0.612	0.798	0.227
Vitamin B6	0.829	0.861	0.833	0.520	0.756
Folate	0.781	0.889	0.889	0.403	0.859
Vitamin B12	-0.023	0.352	0.256	0.154	0.157
Vitamin C	0.254	0.401	0.435	0.331	0.147
Vitamin D	-0.092	0.129	0.150	0.101	0.097
Vitamin E	0.237	0.275	0.341	0.190	0.193
Explained variance%	40.411	41.885	38.984	32.747	13.536

Bold factor loadings used for comparing the nutrient pattern

Table S4. Regression coefficients for fasting glucose for 1 SD increase in the derived nutrient pattern scores among urban black South African women

	Thiamine, starch and folate driven nutrients			Fat and animal protein driven nutrients			Retinol, vitamin b12 and calcium driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	-0.198 (-0.372; -0.024)	0,025	0.011	0.072 (-0.247; 0.104)	0,423	0.000	-0.074 (-0.246; 0.097)	0,395	0.002
M2	-0.139 (-0.414; 0.136)	0,321	0.013	0.048 (-0.160; 0.256)	0,649	0.003	-0.047 (-0.225; 0.131)	0,607	0.011
M3	-0.084 (-0.358; 0.190)	0,547	0.037	0.019 (-0.187; 0.225)	0,859	0.037	-0.039 (-0.215; 0.137)	0,660	0.037
M4	-0.227 (-0.683; 0.230)	0.329	0.062	-0.058 (-0.390; 0.274)	0,732	0.062	-0.058 (-0.270; 0.154)	0,591	0.062

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 :(Adjusted for M2 plus Body Mass Index): M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients Fasting glucose units mmolL⁻¹ = millimoles per litre.

Table S5. Regression coefficients for glycated hemoglobin for 1 SD increase in the derived nutrient pattern scores among urban black South African women

	Thiamine, starch and folate driven nutrients			Fat and animal protein driven nutrients			Retinol, vitamin b12 and calcium driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	-0.060 (-0.160; 0.041)	0,246	0.003	-0.033 (-0.134; 0.068)	0,521	0.001	-0.035 (-0.135; 0.066)	0,500	0.001
M2	-0.033 (-0.192; 0.126)	0,679	0.003	-0.001 (-0.121; 0.119)	0,813	0.003	-0.020 (-0.125; 0.085)	0,703	0.003
M3	0.017 (-0.139; 0.173)	0,833	0.054	-0.028 (-0.146; 0.089)	0,635	0.037	-0.018 (-0.120; 0.085)	0,733	0.054
M4	-0.001 (-0.264; 0.263)	0.994	0.106	-0.018 (-0.209; 0.174)	0,854	0.106	-0.015 (-0.139; 0.109)	0,807	0.106

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 :(Adjusted for M2 plus Body Mass Index): M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients Glycated haemoglobin unit=%

Table S6. Regression coefficients for fasting glucose for 1 SD increase in the derived nutrient pattern scores among urban black South African men

	Thiamine, zinc and plant protein driven nutrients			Fat and animal protein driven nutrients			Beta carotene and vitamin c driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	-0.123 (-0.345; 0.100)	0,279	0.004	0.039 (-0.187; 0.264)	0,737	0.000	0.018 (-0.207; 0.244)	0,873	0.000
M2	-0.159 (-0.541; 0.222)	0,412	0.004	0.098 (-0.157; 0.353)	0,449	0.004	0.044 (-0.193; 0.281)	0,714	0.002
M3	-0.069 (-0.453; 0.316)	0,726	0.025	0.057 (-0.198; 0.311)	0,662	0.025	0.018 (-0.218; 0.253)	0,883	0.025
M4	-0.080 (-0.844; 0.685)	0.838	0.087	0.105 (-0.368; 0.578)	0,662	0.087	-0.096 (-0.438; 0.246)	0,582	0.087

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 :(Adjusted for M2 plus Body Mass Index): M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients Fasting glucose units mmolL⁻¹ = millimoles per litre.

Table S7. Regression coefficients for glycated hemoglobin for 1 SD increase in the derived nutrient pattern scores among urban black South African men

	Thiamine, zinc and plant protein driven nutrients			Fat and animal protein driven nutrients			Beta carotene and vitamin c driven nutrients		
	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²	B (95% CI)	p value	R ²
M1	0.045 (-0.041; 0.131)	0,303	0.003	-0.001 (-0.088; 0.085)	0,973	0.000	-0.003 (-0.089; 0.083)	0,937	0.000
M2	0.077 (-0.070; 0.224)	0,301	0.005	-0.018 (-0.116; 0.081)	0,724	0.002	-0.012 (-0.102; 0.078)	0,789	0.001
M3	0.119 (-0.029; 0.266)	0,166	0.030	-0.035 (-0.133; 0.063)	0,483	0.024	-0.022 (-0.112; 0.067)	0,622	0.024
M4	0.179 (-0.116; 0.473)	0.233	0.094	0.054 (-0.128; 0.235)	0,559	0.094	-0.005 (-0.136; 0.126)	0,938	0.094

M1: (crude); M2: (adjusted for M1 plus Log Total Energy), M3 :(Adjusted for M2 plus Body Mass Index): M4: (Adjusted for M3 plus age, smoking, physical activity, alcohol intake, seasonality, education level, PC1, PC2 and PC3); M1= model 1; M2=model 2; M3=model 3; M4= model 4; PC1= Magnesium, phosphorus and plant protein driven nutrients; PC2 = Fat and animal protein driven nutrients; PC3= Starch, dietary fibre and B vitamin driven nutrients Glycated haemoglobin unit=%

CHAPTER 6

Concluding Remarks

The aim of the study was to determine the genetic and dietary determinants of T2D in a black South African population of Setswana descent. This broader aim was split into three specific objectives: 1) to determine the joint association of European and Asian derived genetic variants associated with T2D in the black South African population; 2) to determine population specific rare and common variants from whole genome sequence data associated with T2D risk; 3) to determine the nutrient patterns associated with T2D susceptibility. These objectives are discussed in the following sections by focusing on the summary, limitations and implications of the findings of the study. The recommendations for future research are indicated at the end of the chapter.

6.1 JOINT ASSOCIATION OF EUROPEAN AND ASIAN DERIVED VARIANTS WITH T2D IN A BLACK SOUTH AFRICAN POPULATION

The polygenic risk scores were reported to modestly improve T2D risk prediction in Caucasian and Asian ethnicities (Wang *et al.*, 2015). However, the utility of the polygenic risk scores of variants discovered from Europeans and Asians in the black South African population is not known. Our study is likely to be the first in Africa to report on the polygenic risk scores of T2D associated variants and their clinical utility in an African population group.

We saw that the GRSt, which was made up of all 66 variants, was not associated with T2D. This was contrary to what had been noted by Vassy *et al.* (2014), that polygenic risk scores were associated similarly in populations which were genetically diverse, such as the Caucasians and African Americans (Vassy *et al.*, 2014). We observed that the GRSt which was made up all the 66 variants was not associated with T2D as was expected. However, the GRSn, which comprised of the variants that were significantly associated with T2D in this population, was significantly associated with T2D. We suspect that the great genetic diversity and difference in LD patterns of the African populations compared

with other ethnicities is the reason for the lack of association of GRSt (Teo *et al.*, 2010). This is also in line with the observation that the variants associated with T2D exhibit great ethnic variation (Chen *et al.*, 2012). We noted, however, that the polygenic risk score was more predictive of T2D risk among participants of less than 50 years old and also among those who were lean, (BMI < 25), as had also been reported in earlier studies (de Miguel-Yanes *et al.*, 2011). Although the GRS was of less clinical utility as it only modestly improved by 2% the prediction of T2D risk in addition to other known risk factors, its improved predictivity among the young and lean, who have not yet developed risk factors for T2D, makes it a very attractive approach for early screening initiatives for T2D. This finding is pivotal in that T2D is preventable and early screening initiatives will help lessen the proportion of people diagnosed of T2D and before they have developed irreversible complications. Therefore, future research which will identify variants that are population specific for black South Africans will help in the identification and screening of people susceptible to T2D early enough for them to be enrolled into T2D prevention programmes focusing on physical activity and diet. This will contribute to the reduction of the incidence of T2D among black South Africans.

6.2 NUTRIENT PATTERNS AND T2D SUSCEPTIBILITY

The genetic profile of people is deemed to be stable, however changes in the eating patterns of people who are predisposed to T2D are believed to be fuelling the increase in the prevalence of this disease (Franks *et al.*, 2013). Notably, it is postulated that the shift in dietary patterns from traditional foods to Western foods as a result of urbanisation is contributing to the increased prevalence of NCDs such as T2D (Vorster *et al.*, 2011). However, the coexistence of improvements in fruit and vegetable intake and the adoption of Western eating, habits that can both decrease and increase T2D risk respectively, was reported in urban black South African residents (Hattingh *et al.*, 2008). In order to capture these contrasting dietary aspects and evaluate their combined effect on T2D risk we used nutrient pattern analysis based on principal component analysis, which enables whole diet associations with biomarkers to be determined (Pisa *et al.*, 2015).

We noted that plant driven nutrient patterns were consumed the most and associated significantly with glycated haemoglobin and fasting glucose. Particularly among rural women, the nutrient pattern driven by starch, dietary fibre and B vitamins was significantly associated with lower fasting glucose levels ($\beta = -0.224$ (-0.446; -0.002); $p = 0.048$) and glycated haemoglobin levels ($\beta = -0.162$ (-0.291; -0.033). In rural men the nutrient pattern driven by thiamine, zinc and plant protein was associated with significant reductions in fasting glucose and glycated haemoglobin of ($\beta = -0.435$ (-0.802; -0.067); $p = 0.021$) and ($\beta = -0.301$ (-0.554; -0.049); $p = 0.020$). However, since this population group also ate meat, the consumption of animal protein, fats and sugar was discovered to be higher in the plant driven nutrient patterns of the urban participants and was associated with trends of increases in fasting glucose and glycated haemoglobin levels. However, animal and fat driven nutrient patterns, which are similar to the Western dietary patterns did not have consistent trends of association with the biomarkers of T2D evaluated in our study.

Although there are a limited number of studies regarding the association of nutrient patterns with fasting glucose and glycated haemoglobin, similar findings have been reported in dietary pattern analysis. Dekker *et al.* (2015) reported that although the Western dietary pattern was being adopted by other ethnicities, the consumption of traditional vegetables was retained in these population groups. Thereby, lessening its association with T2D risk (Dekker *et al.*, 2015). Plant based dietary patterns have also been reported to be associated with reductions in T2D risk, as was noted in our study (Sabate & Wien, 2015). However, in view of the small variances in the glycated haemoglobin and fasting glucose that are explained by the multivariate models evaluated, and the cross sectional design of our study, our findings do not indicate a causality perspective. There might also be a possibility that our results might have been affected by residual confounding as a result of the use of QFFQs, which are not a precise measures of actual dietary intake. However, to the best of our knowledge, our study is likely to be the first to evaluate the association of nutrient patterns with fasting and glycated haemoglobin levels among apparently healthy African individuals. Our findings indicate the benefits of the consumption of plant based nutrients on T2D biomarkers. In view of the fact that nutrients are universal in the diets of varied ethnic groups, our results are relevant for multiple ethnicities and might be useful in the prevention of T2D risk.

6.3 EVALUATION OF THE GENETIC DETERMINANTS OF T2D USING NOVEL WHOLE GENOME SEQUECNEES OF SESTWANA PEOPLE

Limited genetic based studies have been conducted among Africans (Ramsay *et al.*, 2011; Ramsay *et al.*, 2016). No GWAS study for T2D has been conducted in Africa and only about 10% of the more than 100 common variants associated with T2D variants are of importance to African ethnicities in Africa (Adeyemo *et al.*, 2015). Added to this, the role of rare variants in the risk of T2D has not been explored in Africa so far. Using an extreme phenotype sampling approach, 16 T2D cases and 14 controls were selected from the 1% tails of the glycated haemoglobin distribution after adjustment for age, sex, BMI and fasting glucose. This enabled the sampling of individuals from the tails of unexplained variation of glycated haemoglobin, which are probably for by genetic factors.

We discovered that about 54% of the variants of the whole genome sequences of the Setswana population were novel, based on the dbSNP version 131 that was used to annotate the variants. However the proportion of this novelty will need to be evaluated using the current dbSNP 148 version. Regardless, this indicated that a significant proportion of the variation among the black South Africans is yet to be captured. We discovered that the rs7499753 SNP, located in the intron of the *CIITA* gene, had an association with T2D that exceeded the genome wide statistical significance. A number of coding variants with nominal significance of $p < 5 \times 10^{-5}$ were noted. However, these SNPs have to be replicated first to validate their association with T2D. The rs7499753 variant was reported to be insignificant in a T2D genetic association study among Asians, Europeans and African Americans (Fuchsberger *et al.*, 2016). However, the discrepancy in the association of rs7499753 with T2D in our study and that of Fuchsberger *et al.*, (2016) can be explained by the differences in the allele frequencies of this SNP we noted among these studies. Differences in allele frequencies have been discovered to lead to the varied associations of SNPs in different ethnicities (Teo *et al.*, 2010). Nonetheless, the association of the rs7499753 SNP with T2D needs to be confirmed by replications studies.

None of the genes with non-synonymous rare variants was able to reach genome wide significance. In view of our small sample sizes, SNPs which have high effect sizes on T2D risk are the ones which might reach genome wide significance; this might explain the high ORs which were noted for the discovered variants (Kang *et al.*, 2012). However, since the logistic ORs have been observed to be inflated in studies with small sample sizes, more well powered studies will be required to validate them (Nemes *et al.*, 2009; Button *et al.*, 2013). Regardless that extreme phenotype sampling is known to improve the power to detect rare variants, our sample size might have low statistical power to detect these variants and better powered studies are required to address this phenomenon (Guey *et al.*, 2011). Nonetheless, our study, to the best of our knowledge, has pioneered the use of whole genome sequences to determine genetic variants associated with T2D risk in Africa. It will therefore serve as a guide to how similar studies in this domain can be conducted.

6.4 IMPLICATIONS OF FUTURE RESEARCH AND RECOMMENDATIONS

The findings of this PhD study have contributed to the knowledge base of the genetic and dietary determinants of T2D among the Setswana black South Africans. We discovered the following in this study: 1) Polygenic risk scores are more predictive of T2D risk when they consists of variants that are significantly associated with T2D in the Setswana black South African population. These polygenic risk scores are more predictive of T2D risk among the lean and those less than 50years of age; 2) A possible association of rs7499753 with T2D was observed which needs to be validated with replication studies. The Setswana were observed to share common ancestries with the Zulu and Sotho ethnolinguistic groups. Thus the SNPs associated with T2D in the Setswana might be relevant in the Zulu and Sotho ethnolinguistic groups. 3) The consumption of plant driven nutrient patterns was associated with reductions in fasting and glycated hemoglobin. However, additional questions for further investigation have arisen which might help to unravel more of the dietary and genetic determinants of this disease. These questions were elaborated on in the following sections.

6.4.1 Polygenetic risk scores and T2D

The addition of more population specific rare and common variants to the GRS through whole genome association studies and GWASs is required to help improve predictivity of T2D risk in African populations. Our study was cross sectional in nature and it will be pivotal to evaluate the GRS longitudinal associations using net reclassification approaches that have been noted to be more sensitive than the ROC used in this study to evaluate the predictive utility of the polygenic risk scores (Wang *et al.*, 2015). More studies will also be required to develop these GRSs based on pathways relevant to T2D, which has a heterogeneous aetiology, in order to determine more effective GRSs that predict T2D risk in varied settings.

6.4.2 Nutrient patterns and T2D

The nutrient pattern associations with fasting glucose and glycated haemoglobin were evaluated in a cross sectional design. It will be pivotal to assess these associations in longitudinal study designs that are less biased. QFFQs used in this study have been associated with residual confounding and the adjustment of energy intake has been noted to be incomplete (Willett, 1998). Therefore clinical trials using 24-hour recalls and iso-caloric diets are required to determine the nutrient pattern causality associated with fasting and glycated haemoglobin levels.

6.4.3 Rare variants and T2D risk

The extreme phenotype sampling used in this study is known to boost power to detect rare variants. However, Kang *et al* (2012) indicated the benefits of a two stage extreme phenotype sampling design, which was not implemented in our study. He indicates that, with as little as 0.2% of the study population selected from the extreme tails in the first stage, about 80% of the causal rare variants can be detected, though they might not reach genome wide significance, and will be discovered upon replication in the second stage of the study (Kang *et al.*, 2012). This approach and the validation of the top signals in the single variant tests are thus pivotal in the future. Although the AVGP was a collaborative

whole genome sequencing project in Africa, it was focused mostly on documenting the genetic variation; large, disease-based whole genome sequencing collaborations are now required to unravel the genetic determinants of T2D in Africa.

6.5 REFERENCES

Adeyemo, A.A., Tekola-Ayele, F., Doumatey, A.P., Bentley, A.R., Chen, G., Huang, H., Zhou, J., Shriner, D., Fasanmade, O., Okafor, G., Eghan, B., Jr., Agyenim-Boateng, K., Adeleye, J., Balogun, W., Elkahloun, A., Chandrasekharappa, S., Owusu, S., Amoah, A., Acheampong, J., Johnson, T., Oli, J., Adebamowo, C., Collins, F., Dunston, G. & Rotimi, C.N. 2015. Evaluation of Genome Wide Association Study Associated Type 2 Diabetes Susceptibility Loci in Sub Saharan Africans. *Front Genet*, 6:335.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J. & Munafo, M.R. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*, 14(5):365-376.

Chen, R., Corona, E., Sikora, M., Dudley, J.T., Morgan, A.A., Moreno-Estrada, A., Nilsen, G.B., Ruau, D., Lincoln, S.E., Bustamante, C.D. & Butte, A.J. 2012. Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet*, 8(4):e1002621.

de Miguel-Yanes, J.M., Shrader, P., Pencina, M.J., Fox, C.S., Manning, A.K., Grant, R.W., Dupuis, J., Florez, J.C., D'Agostino, R.B., Sr., Cupples, L.A., Meigs, J.B., Investigators, M. & Investigators, D. 2011. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes Care*, 34(1):121-125.

Dekker, L.H., van Dam, R.M., Snijder, M.B., Peters, R.J., Dekker, J.M., de Vries, J.H., de Boer, E.J., Schulze, M.B., Stronks, K. & Nicolaou, M. 2015. Comparable Dietary Patterns Describe Dietary Behavior across Ethnic Groups in the Netherlands, but Different Elements in the Diet Are Associated with Glycated Hemoglobin and Fasting Glucose Concentrations. *J Nutr*, 145(8):1884-1891.

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.

Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., Rivas, M.A., Perry, J.R.B., Sim, X., Blackwell, T.W., Robertson, N.R., Rayner, N.W., Cingolani, P., Locke, A.E., Tajes, J.F., Highland, H.M., Dupuis, J., Chines, P.S., Lindgren, C.M., Hartl, C., Jackson, A.U., Chen, H., Huyghe, J.R., van de Bunt, M., Pearson, R.D., Kumar, A., Muller-Nurasyid, M., Grarup, N., Stringham, H.M., Gamazon, E.R., Lee, J., Chen, Y.H., Scott, R.A., Below, J.E., Chen, P., Huang, J., Go, M.J., Stitzel, M.L., Pasko, D., Parker, S.C.J., Varga, T.V., Green, T., Beer, N.L., Day-Williams, A.G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M.K., Kim, B.J., Kim, Y., Kim, Y.J., Kwon, M.S., Lee, J., Lee, S., Lin,

K.H., Maxwell, T.J., Nagai, Y., Wang, X., Welch, R.P., Yoon, J., Zhang, W., Barzilai, N., Voight, B.F., Han, B.G., Jenkinson, C.P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M.C.Y., Palmer, N.D., Balkau, B., Stancakova, A., Abboud, H.E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J.D., Neale, B.M., Purcell, S., Butterworth, A.S., Howson, J.M.M., Lee, H.M., Lu, Y.C., Kwak, S.H., Zhao, W., Danesh, J., Lam, V.K.L., Park, K.S., Saleheen, D., So, W.Y., Tam, C.H.T., Afzal, U., Aguilar, D., Arya, R., Aung, T., Chan, E., Navarro, C., Cheng, C.Y., Palli, D., Correa, A., Curran, J.E., Rybin, D., Farook, V.S., Fowler, S.P., Freedman, B.I., Griswold, M., Hale, D.E., Hicks, P.J., Khor, C.C., Kumar, S., Lehne, B., Thuillier, D., Lim, W.Y., Liu, J., van der Schouw, Y.T., Loh, M., Musani, S.K., Puppala, S., Scott, W.R., Yengo, L., Tan, S.T., Taylor, H.A., Thameem, F., Wilson, G., Wong, T.Y., Njolstad, P.R., Levy, J.C., Mangino, M., Bonnycastle, L.L., Schwarzmayr, T., Fadista, J., Surdulescu, G.L., Herder, C., Groves, C.J., Wieland, T., Bork-Jensen, J., Brandslund, I., Christensen, C., Koistinen, H.A., Doney, A.S.F., Kinnunen, L., Esko, T., Farmer, A.J., Hakaste, L., Hodgkiss, D., Kravic, J., Lyssenko, V., Hollensted, M., Jorgensen, M.E., Jorgensen, T., Ladenvall, C., Justesen, J.M., Karajamaki, A., Kriebel, J., Rathmann, W., Lannfelt, L., Lauritzen, T., Narisu, N., Linneberg, A., Melander, O., Milani, L., Neville, M., Orho-Melander, M., Qi, L., Qi, Q.B., Roden, M., Rolandsson, O., Swift, A., Rosengren, A.H., Stirrups, K., Wood, A.R., Mihailov, E., Blancher, C., Carneiro, M.O., Maguire, J., Poplin, R., Shakir, K., Fennell, T., DePristo, M., de Angelis, M.H., Deloukas, P., Gjesing, A.P., Jun, G., Nilsson, P., Murphy, J., Onofrio, R., Thorand, B., Hansen, T., Meisinger, C., Hu, F.B., Isomaa, B., Karpe, F., Liang, L.M., Peters, A., Huth, C., O'Rahilly, S.P., Palmer, C.N.A., Pedersen, O., Rauramaa, R., Tuomilehto, J., Salomaa, V., Watanabe, R.M., Syvanen, A.C., Bergman, R.N., Bharadwaj, D., Bottinger, E.P., Cho, Y.S., Chandak, G.R., Chan, J.C.N., Chia, K.S., Daly, M.J., Ebrahim, S.B., Langenberg, C., Elliott, P., Jablonski, K.A., Lehman, D.M., Jia, W.P., Ma, R.C.W., Pollin, T.I., Sandhu, M., Tandon, N., Froguel, P., Barroso, I., Teo, Y.Y., Zeggini, E., Loos, R.J.F., Small, K.S., Ried, J.S., DeFronzo, R.A., Grallert, H., Glaser, B., Metspalu, A., Wareham, N.J., Walker, M., Banks, E., Gieger, C., Ingelsson, E., Im, H.K., Illig, T., Franks, P.W., Buck, G., Trakalo, J., Buck, D., Prokopenko, I., Magi, R., Lind, L., Farjoun, Y., Owen, K.R., Gloyn, A.L., Strauch, K., Tuomi, T., Kooner, J.S., Lee, J.Y., Park, T., Donnelly, P., Morris, A.D., Hattersley, A.T., Bowden, D.W., Collins, F.S., Atzmon, G., Chambers, J.C., Spector, T.D., Laakso, M., Strom, T.M., Bell, G.I., Blangero, J., Duggirala, R., Tai, E.S., McVean, G., Hanis, C.L., Wilson, J.G., Seielstad, M., Frayling, T.M., Meigs, J.B., Cox, N.J., Sladek, R., Lander, E.S., Gabriel, S., Burt, N.P., Mohlke, K.L., Meitinger, T., Groop, L., Abecasis, G., Florez, J.C., Scott, L.J., Morris, A.P., Kang, H.M., Boehnke, M., Altshuler, D. & McCarthy, M.I. 2016. The genetic architecture of type 2 diabetes. *Nature*, 536(7614):41-+.

Guey, L.T., Kravic, J., Melander, O., Burt, N.P., Laramie, J.M., Lyssenko, V., Jonsson, A., Lindholm, E., Tuomi, T., Isomaa, B., Nilsson, P., Almgren, P., Kathiresan, S., Groop, L., Seymour, A.B., Altshuler, D. & Voight, B.F. 2011. Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet Epidemiol*, 35(4):236-246.

Hattingh, Z., Walsh, C.M., Bester, C.J. & Oguntibeju, O.O. 2008. Evaluation of energy and macronutrient intake of black women in Bloemfontein: A cross-sectional study. *African Journal of Biotechnology*, 7(22):4019-4024.

- Kang, G., Lin, D., Hakonarson, H. & Chen, J. 2012. Two-stage extreme phenotype sequencing design for discovering and testing common and rare genetic variants: efficiency and power. *Hum Hered*, 73(3):139-147.
- Nemes, S., Jonasson, J.M., Genell, A. & Steineck, G. 2009. Bias in odds ratios by logistic regression modelling and sample size. *BMC Medical Research Methodology*, 9(1):56.
- Pisa, P.T., Pedro, T.M., Kahn, K., Tollman, S.M., Pettifor, J.M. & Norris, S.A. 2015. Nutrient patterns and their association with socio-demographic, lifestyle factors and obesity risk in rural South African adolescents. *Nutrients*, 7(5):3464-3482.
- Ramsay, M., Sankoh, O., as members of the, A.W.I.G.s. & the, H.A.C. 2016. African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent. *Int J Epidemiol*, 45(2):305-308.
- Ramsay, M., Tiemessen, C.T., Choudhury, A. & Soodyall, H. 2011. Africa: the next frontier for human disease gene discovery? *Human molecular genetics*, 20(R2):R214-220.
- Sabate, J. & Wien, M. 2015. A perspective on vegetarian dietary patterns and risk of metabolic syndrome. *Br J Nutr*, 113 Suppl 2:S136-143.
- Teo, Y.Y., Small, K.S. & Kwiatkowski, D.P. 2010. Methodological challenges of genome-wide association analysis in Africa. *Nature reviews.Genetics*, 11(2):149-160.
- Vassy, J.L., Hivert, M.F., Porneala, B., Dauriz, M., Florez, J.C., Dupuis, J., Siscovick, D.S., Fornage, M., Rasmussen-Torvik, L.J., Bouchard, C. & Meigs, J.B. 2014. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes*, 63(6):2172-2182.
- Vorster, H.H., Kruger, A. & Margetts, B.M. 2011. The Nutrition Transition in Africa: Can It Be Steered into a More Positive Direction? *Nutrients*, 3(4):429-441.
- Wang, X., Strizich, G., Hu, Y., Wang, T., Kaplan, R.C. & Qi, Q. 2015. Genetic markers of type 2 diabetes: Progress in genome-wide association studies and clinical application for risk prediction. *J Diabetes*.
- Willett, W. 1998. Nutritional epidemiology. New York: Oxford University Press.

APPENDIX A

Ethical Approval and Consent forms



NORTH-WEST UNIVERSITY
YUNIBESITHI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOM CAMPUS

Private Bag X6001, Potchefstroom
South Africa 2520

Tel: 018 299-1111/2222
Web: <http://www.nwu.ac.za>

Ethics Office
Tel: 018-285 2291

Email: Wayne.Tovers@nwu.ac.za

15 February 2016

Dr KR Conradie
CEN

Dear Dr Conradie

HREC APPROVAL OF YOUR APPLICATION

Ethics number: NWU-00016-10-A1 PURE Study

Kindly use the ethics reference number provided above in all correspondence or documents submitted to the Health Research Ethics Committee (HREC) secretariat.

Project title: Evaluation of gene-diet interactions associated with type 2 diabetes susceptibility in a black South African population

Project leader/supervisor: Dr KR Conradie

Student: T Chikowore

Application type: Sub-study

Risk level descriptor: Medium

You are kindly informed that at the meeting held on 15/07/2015 of the HREC, Faculty of Health Sciences, the aforementioned was approved.

The period of approval for this project is from 15/02/2016 to 31/12/2016.

After ethical review:

Translation of the informed consent document to the languages applicable to the study participants should be submitted to the HREC (if applicable).

The HREC requires immediate reporting of any aspects that warrants a change of ethical approval. Any amendments, extensions or other modifications to the protocol or other associated documentation must be submitted to the HREC prior to implementing these changes. Any adverse/unexpected/unforeseen events or incidents must be reported on either an adverse event report form or incident report form.

A progress report should be submitted within one year of approval of this study and before the year has expired, to ensure timely renewal of the study. A final report must be provided at completion of the study or the HREC must be notified if the study is temporarily suspended or terminated. The progress report template is obtainable from Carolien van Zyl at

Carolien.VanZyl@nwu.ac.za. Annually a number of projects may be randomly selected for an external audit.

Please note that the HREC has the prerogative and authority to ask further questions, seek additional information, require further modification or monitor the conduct of your research or the informed consent process.

Please note that for any research at governmental or private institutions, permission must still be obtained from relevant authorities and provided to the HREC. Ethics approval is required BEFORE approval can be obtained from these authorities.

The HREC complies with the South African National Health Act 61 (2003), the regulations on Research with Human Participants of 2014 of the Department of Health and Principles, the Declaration of Helsinki, 2013, the Belmont Report and the Ethics in Health Research: Principles, Structures and Processes (SANS document).

We wish you the best as you conduct your research. If you have any questions or need further assistance, please contact the Ethics Office at Carolien.VanZyl@nwu.ac.za or 018 299 1206.

Yours sincerely



Dr Wayne Towers
HREC Chairperson



Prof Minnie Greeff
Ethics Office Head

Current details: (13210572) C:\Users\13210572\Documents\HREC\HREC - Applications\2015 Applications\Applications 06 - 15 July 2015\NWU-00016-10-A1 (KR Conradie-T Chikowore)\NWU-00016-10-A1 (KR Conradie-T Chikowore) - AL\NWU-00016-10-A1 (KR Conradie-T Chikowore) - AL.docm
15 February 2016

File reference: 9.1.5.3

HREC Stamp

**PARTICIPANT INFORMATION LEAFLET AND CONSENT FORM
FOR PARTICIPANTS SELECTED TO BE CASES FOR A
SMALLER STUDY WITHIN THE PURE STUDY.**

**TITLE OF THE RESEARCH PROJECT: EVALUATION OF GENE-DIET
INTERACTIONS ASSOCIATED WITH TYPE 2 DIABETES IN A BLACK SOUTH
AFRICAN POPULATION**

REFERENCE NUMBERS: NWU-00016-10-A1

PRINCIPAL INVESTIGATOR: Dr K Conradie

ADDRESS: 11 Hoffman Street, Potchefstroom

CONTACT NUMBER: +27 (0) 846080878

You are being invited to take part in a smaller research project that forms part of the bigger PURE study. Please take some time to read the information presented here, which will explain the details of this project. Please ask the researcher any questions about any part of this project that you do not fully understand. It is very important that you are fully satisfied that you clearly understand what this research entails and how you could be involved. Also, your participation is **entirely voluntary** and you are free to decline to participate. If you say no, this will not affect you negatively in any way whatsoever. You are also free to withdraw from the study at any point, even if you do agree to take part.

This study has been approved by the **Health Research Ethics Committee of the Faculty of Health Sciences of the North-West University (NWU. NWU-00016-10-A1)** and will be conducted according to the ethical guidelines and principles of the international Declaration of Helsinki and the ethical guidelines of the National Health Research Ethics Council. This research forms part of the PhD students study. It might
HREC General WICF Version 3, March 2015

Page 1 of 5

be necessary for the research ethics committee members or relevant authorities to inspect the research records.

What is this research study all about?

- *The study aims to assess the full set of instructions which is found in your blood called your genetic material or your DNA. You got this set of instructions from both your parents. This DNA is like a manual that tells your body how to work. Changes can occur in this manual (DNA) over time, which can lead to your body not working right and getting a disease like high blood sugar (diabetes). These changes are common in certain families. That is why certain families have people who are more likely to have high blood sugar (diabetes). We would like to search for the changes in your manual (DNA) which might lead to the high blood sugar which is found in adults.*
- *Since your blood was collected in 2005 you are being asked to give approval for us to take out the manual (DNA) from your blood to determine the changes that might cause you to get high blood sugar (diabetes). To do this we have to look at all 3 billion letters in the manual. These are called nucleotides and we can read them using a machine called a next generation sequencer.*

Why have you been invited to participate?

- *You have been invited to participate because you gave permission for your blood to be drawn in 2005 and be analysed by the trained researchers at the North-West University.*
- *You also have one of the following inclusion criteria: You have either some history of high blood sugar (diabetes) in your family, or you have the high blood sugar (diabetes) or high you are at high risk of developing high blood sugar (diabetes) due to high levels of markers for blood sugar.*

What will your responsibilities be?

- *You will be asked to consent for the analysis of your manual (DNA) from the blood you gave for analysis in 2005. We also want to ask that you allow us to send the DNA overseas to be analysed if we fail to find local machines which are used to analyse it in South Africa.*

Will you benefit from taking part in this research?

- *There will be no immediate benefit to you yourself. However through your participation in this study, important information will be obtained on the changes in the manual (DNA) of other black South African adults which can help us to understand how high blood sugar (diabetes) develops. This will help us to cure and prevent high blood sugar (diabetes). This will help to identify the disease early in people who might have the high blood sugar (diabetes) disease before they are too sick.*

Are there risks involved in your taking part in this research?

- *There will be no physical risk as we will not take any blood or measurements from you*
- *Possible other risks might be that your manual (DNA) may be used for other purposes other than studying high blood sugar (diabetes) if the analysis is done overseas. However, we have a contract with the people who will do this analysis*

overseas that your manual (DNA) will only be used for this study and then destroyed. All data will be protected by passwords and stored in locked rooms.

- *Researchers are using information from white people to determine how diseases are inherited across families and populations. The genetic material of black people has changes which are different from those of the white people. Therefore it is necessary that research will also be done on black people in South Africa. Therefore this study might help in identifying these differences in the black population which may cause people to be at risk of getting high blood sugar (diabetes) when they are adults. Because the risk of your data being used by other people is minimal, the benefit of having medication and tests to help black people in the future to not get the sugar disease is much higher than the risk involved in this study.*
- *By looking into your manual (DNA) we might find also how you are at risk of developing other diseases besides high sugar blood. If you would like us to report back your risk of other diseases you may indicate so in the declaration section.*

What will happen in the unlikely event of some form of discomfort occurring as a direct result of your taking part in this research study?

- *Should you have the need for further discussions after you experience any discomfort as a result of this study an opportunity will be arranged for you to raise it up to the researches and the NWU Health research ethics committee.*

Who will have access to the data?

- *Your information will be confidential and unidentifiable and will only be identifiable to the principal investigator (PI) of the PURE study by means of a unique code. Therefore anonymity will be ensured. Reporting of findings will also be anonymous. The findings will only be describing the results as generic from a black South African population and not even the unique codes will be given. Only the researchers will have access to the codified data. Data will be kept safe and secure by locking hard copies in locked cupboards in the researcher's office and for electronic data it will be password protected. Data will be stored for 5 years.*

What will happen with the data/samples?

- *This genetic material will be destroyed by the company performing the analysis to avoid them being used for other purposes besides those you would have consented to.*

Will you be paid to take part in this study and are there any costs involved?

You will not be paid to take part in the study as it will not cost you anything. The researchers and the field officers will visit you and explain the study to you. You will be given time to read through this form and then approve to be part of this study by means of a signature.

Is there anything else that you should know or do?

- *You can contact Dr K Conradie at +27 (0) 846080878 if you have any further queries or encounter any problems.*

- You can contact the Health Research Ethics Committee via Mrs Carolien van Zyl at 018 299 2089; carolien.vanzyl@nwu.ac.za if you have any concerns or complaints that have not been adequately addressed by the researcher.
- You will receive a copy of this information and consent form for your own records.

How will you know about the findings?

- The findings of the research will be shared with you after they have been thoroughly evaluated and similar results are shown also in a larger study. Meetings will be scheduled in consultation with the PURE study leaders to report back to you and other participants about the results from this study.

Declaration by participant

By signing below, I agree to take part in a research study titled: Evaluation of gene diet interactions associated with type 2 diabetes in a black South African population.

I declare that:

- I have read this information and consent form and it is written in a language with which I am fluent and comfortable.
- I have had a chance to ask questions to both the person obtaining consent, as well as the researcher and all my questions have been adequately answered.
- I understand that taking part in this study is **voluntary** and I have not been pressurised to take part.
- I may choose to leave the study at any time and will not be penalised or prejudiced in any way.
- I may be asked to leave the study before it has finished, if the researcher feels it is in my best interests, or if I do not follow the study plan, as agreed to.
- I understand that the full set of the instructions in my blood which governs how the body functions will be analysed overseas by means of whole genome sequencing.
- I would like information regarding changes in my DNA other than that related to high sugar (diabetes) to be reported back to me. (Indicate by a Y for yes and N for No in the text box). ☐

Signed at (*place*) on (*date*) 20....

.....
Signature of participant

.....
Signature of witness

Declaration by person obtaining consent

I (name) declare that:

- I explained the information in this document to
- I encouraged him/her to ask questions and took adequate time to answer them.
- I am satisfied that he/she adequately understands all aspects of the research, as discussed above
- I did/did not use a interpreter.

Signed at (place) on (date) 20....

.....
Signature of person obtaining consent

.....
Signature of witness

Declaration by researcher

I (name) declare that:

- I explained the information in this document to
- I encouraged him/her to ask questions and took adequate time to answer them.
- I am satisfied that he/she adequately understands all aspects of the research, as discussed above
- I did/did not use a interpreter.

Signed at (place) on (date) 20....

.....
Signature of researcher

.....
Signature of witness

APPENDIX B

Published Manuscript



Contents available at ScienceDirect

Diabetes Research
and Clinical Practicejournal homepage: www.elsevier.com/locate/diabresInternational
Diabetes
Federation

Predictive utility of a genetic risk score of common variants associated with type 2 diabetes in a black South African population

Tinashe Chikowore^{a,*}, Tertia van Zyl^a, Edith J.M. Feskens^b, Karin R. Conradie^a

^a Centre for Excellence in Nutrition, North-West University, Potchefstroom, North West Province 2520, South Africa

^b Wageningen University, Division of Human Nutrition, P.O. Box 17, 6700 AA Wageningen, The Netherlands

ARTICLE INFO

Article history:

Received 2 May 2016

Received in revised form

21 September 2016

Accepted 22 September 2016

Available online 1 October 2016

Keywords:

GRS

Polygenic risk score

Clinical use

BMI

ABSTRACT

Aims: To determine the predictive utility of polygenic risk scores of common variants associated with type 2 diabetes derived from the European and Asian ethnicities among a black South African population.

Method: Our study was a case-control study nested within the Prospective Urban and Rural Epidemiological (PURE) study of 178 male and female cases, matched for age and gender with 178 controls. Four types of genetic risk scores (GRS) were developed from 66 selected SNPs. These comprised of beta cell related variants (GRSb), variants which had significant associations with T2D in our study (GRSn), variants from the trans-ethnic meta-analysis (GRStrans) and all the 66 selected SNPs (GRSt).

Results: Of the GRS's, only GRSn was associated with increased risk of T2D as indicated by an OR (95CI) of 1.21 (1.02–1.43) p -value = 0.015. Stratified analysis of age and BMI, indicated the GRSn to be significantly associated with T2D among the non-obese and participants less than 50 years. The area under the ROC of the T2D risk factors only was 0.652 (p value < 0.001) and with the addition of GRSn it was 0.665 (p value < 0.001).

Conclusions: The GRS of European and Asian derived variants have limited clinical utility in the black South African population. The inclusion of population specific variants in the GRS is pivotal.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The prevalence of the type 2 diabetes (T2D) disease burden is increasing globally and South Africans have not been spared [1,2]. Approximately 2 million South Africans live with T2D, with about 115,000 new cases occurring annually [1]. However, limited evidence exists about the etiology of T2D, which is hampering the development of effective therapeutic and preventative interventions to curb this growing burden [3].

Added to this, there is an urgent need to scale up diabetes screening, as it is estimated that 30–80% of the new T2D cases remain undiagnosed and that even in those that are diagnosed, 20% of them would have developed irreversible complications by the time of diagnoses [4].

The strong genetic component and heritability of T2D which is indicated by concordance rates of 79% among monozygotic twins provides an opportunity to identify high risk individuals well before they have developed the disease

* Corresponding author. Fax: +27 182992464.

E-mail address: tinashedoc@gmail.com (T. Chikowore).

<http://dx.doi.org/10.1016/j.diabres.2016.09.019>

0168-8227/© 2016 Elsevier Ireland Ltd. All rights reserved.

[5,6]. The genetic risk score (GRS) is an approach that combines multiple SNP information into a single metric representing the genetic predisposition for an individual to a particular phenotype or disease trait [7]. The GRS has been depicted to be more predictive in participants who are non-obese and not older than 50 years of age [8]. Using the GRS to identify high risk individuals while still young and without risk factors such as obesity being present, is therefore very relevant. Although individual variants have been depicted to have small effect sizes, there is a growing interest that a combination of these variants in the form of a GRS might be more useful in predicting disease susceptibility [9]. As opposed to the genome wide association studies (GWAS) which require large sample sizes to identify significant variants, the GRS were noted not to vary as a result of sample size and study design [10]. Thereby, indicating the GRS approach relevant for evaluating genetic risk in modest samples sizes that are common among study settings in Africa due to resource constraints.

Numerous studies have been done among Caucasians to evaluate the effectiveness of the GRS alone and in combination with conventional T2D risk factors such as age, sex and body mass index (BMI) [7,9,10]. However, it has been noted to only modestly improve the predictive ability of the conventional risk factors of T2D among the Caucasians. Little is known of the association of GRS of common variants derived from European and Asian ethnicities with T2D in Africa [7,10]. However, the GRS of European derived common variants was reported to be significantly associated with T2D susceptibility among Caucasians and African Americans regardless that individual common variants associated with T2D in these ethnic groups are different [9]. Linkage disequilibrium (LD) differences among different ethnic groups have been identified as the main factor for not being able to replicate the common variants associations [11]. However, it seems that the GRS is not limited by LD and this aspect needs to be further investigated in other populations of African ancestry which have not been able to replicate the association of common variants with T2D [9]. This study was done to determine the GRS constructed from common variants derived from European and Asian ethnicities associated with T2D, its predictive utility alone and in combination with other conventional risk factors in a black South African population of Setswana descent.

2. Material and methods

2.1. Study participants

One hundred and seventy eighty cases and 178 controls were recruited for this study from the South African arm of the PURE study as described elsewhere [12]. The 178 cases were matched for gender and age with 178 controls. The cases were identified as having blood glucose levels above 7.8 mmol L^{-1} while the controls had normal glucose levels (less than 7.8 mmol L^{-1}) after a two-hour oral glucose tolerance test (OGTT) were administered to each participant. The participants completed written informed consents to participate in the study. The study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval

was obtained from the Ethics Committee of the North-West University, Potchefstroom Campus, with ethics number NWU-00016-10-S1.

2.2. Biochemical measurements

The research participants who were in a fasting state (at least 10 h with no food or beverages, including water) had their blood glucose levels measured by the PURE research team. The fasting glucose levels were measured using the SYN-CHRON® System from fluoride plasma. Thereafter, the OGTT was performed by giving the participants a 75 g glucose load dissolved in 300 mL of water to ingest and post-prandial glucose measurement from were determined from withdrawn blood at 30, 60, 90 and 120 min as described in detail elsewhere [12]. The Bio-Rad D-10™, HbA1c kit (Bio-Rad Laboratories, Inc., Hercules, France), which operates via cation exchange high performance liquid chromatography was used to assess HbA1c levels from whole blood EDTA samples.

2.3. BMI and blood pressure

The height was measured using a stadiometer (IP 1465, Invicta, and London, UK) and weight was assessed using portable electronic scales (Precision Health Scale, A&D Tokyo, Japan). The measurements were done in duplicate according to standardised protocols and the mean was recorded. BMI was computed as, $\text{BMI} = \text{weight (kilograms)} / \text{height (meters)}^2$. The participants were allowed a 10-min rest period, after which BP was measured twice, after a duration of 5 min. Blood pressure was measured on the upper right arm while the participants were seated upright and the arm supported at heart level. The OMRON HEM-757 (Omron Healthcare, Kyoto, Japan) was used to measure systolic and diastolic blood pressure. Appropriately sized cuffs were used for the obese participants.

2.4. DNA preparation and genotyping

The Qiagen Flexigene DNA extraction kit (catalogue number 51206, QIAGEN Pty. Ltd., Australia) was used to extract genomic DNA from citrated treated buffy coat. Seventy-nine SNPs associated with T2D were selected from literature and genotyped using the GoldenGate Assay with VeraCode technology (Illumina Inc., San Diego, CA, USA) on the BeadXpress platform (Illumina Inc., San Diego, CA, USA). After quality control 66 SNPs remained for association tests. The details regarding the DNA preparation, genotyping and quality control are described in more detail elsewhere [12].

2.5. Construction of the GRS

A simple count (unweighted) method was used to calculate the GRS which assumes the additive genetic model for each SNP. However, the weighted GRS approach which takes into account the different effects sizes of SNPs on T2D risk was not used due to the absence of GWAS studies for T2D in African populations, from which it is recommended to select effect sizes for constructing the weighted GRS [13]. Effect size of the risk alleles could not be selected from GWAS studies

Table 1 – Genetic risk scores and their related variants.

GRS	Nearest loci	SNP ID	Risk allele	Other allele	Study used to designate the SNPs into a selected category
GRSn	FTO	rs8050136	A	C	Chikowore et al. [12]
	PSMD6	rs831571	C	T	
	SLC44A3, F3	rs7542900	C	T	
	C2CD4B	rs1436955	C	T	
GRSbeta	MTNR1B	rs1387153	T	C	Qi and Hu [21]
	TCF7L2	rs7903146	T	C	
	C2CD4B	rs1436955	C	T	
	SLC30A8	rs1326634	T	C	
	CDKAL1	rs7754840	C	A	
GRStrans	HHEX	rs1111875	C	T	Morris et al. [23]
	IGF2BP2	rs4402960	T	G	
	CDKN2A, CD-KN2B	rs564398	T	C	
	PRC1	rs8042680	A	C	
	JAZF1	rs864745	T	C	
	MTNR1B	rs1387153	T	C	
	TCF7L2	rs7903146	T	C	
	SLC30A8	rs1326634	T	C	
	CDKAL1	rs7754840	C	A	
GRSt	LPIN2	rs10460009	C	T	Sim et al. [24]
	C6orf57	rs1048886	G	A	Sim et al. [24]
	TCERG1L	rs10741243	G	C	Sim et al. [24]
	NOTCH2, ADAM30	rs10923931	T	G	Zeggini et al. [25]
	HHEX	rs1111875	C	T	Scott et al. [26]
	DCD	rs1153188	A	T	Zeggini et al. [25]
	ZFAND6	rs11634397	A	G	Voight et al. [27]
	LOC72901, CETN3	rs12518099	G	A	Rung et al. [28]
	CDC123, CAMK1D	rs12779790	G	A	Zeggini et al. [25]
	SLC30A8	rs1326634	T	C	Scott et al. [26]
	CHCHD9	rs13292136	C	T	Voight et al. [27]
	SPRY2	rs1359790	G	A	Shu et al. [29]
	MTNR1B	rs1387153	T	C	Voight et al. [27]
	C2CD4B	rs1436955	C	T	Shu et al. [29]
	HMGA2	rs1531343	C	G	Voight et al. [27]
	ST6GAL1	rs16861329	C	T	Kooner et al. [30]
	CMIP	rs16955379	C	T	Cho et al. [31]
	SYN2, PPARG	rs17036101	G	A	Zeggini et al. [25]
	PTPRD	rs17584499	T	C	Tsai et al. [32]
	WVVOX	rs17797882	T	C	Cho et al. [31]
	AP3S2	rs2028299	C	A	Kooner et al. [30]
	KCNQ1	rs231362	G	A	Voight et al. [27]
	BCL11A	rs243021	A	G	Voight et al. [27]
	GALNTL4, LOC729013	rs2722769	C	G	Palmer et al. [33]
	HUNK	rs2833610	A	G	Sim et al. [24]
	LOC64673, IRS1	rs2943641	C	T	Rung et al. [28]
	NR	rs358806	A	C	WTCC Consortium, 2007
	PLS1	rs3773506	C	G	Sim et al. [24]
	SRR	rs391300	C	T	Tsai et al. [32]
	GRB14	rs3923113	A	C	Kooner et al. [30]
	IGF2BP2	rs4402960	T	G	Scott et al. [26]
	HNF1B	rs4430796	G	A	Voight et al. [27]
	ZBED3	rs4457053	G	A	Voight et al. [27]
	ADAMTS9	rs4607103	C	T	Zeggini et al. [25]
	WFS1, PPP2R2C	rs4689388	A	G	Rung et al. [28]
	PAPL	rs472265	G	A	Sim et al. [24]
	HNF4A	rs4812829	A	G	Kooner et al. [30]
	CDKN2A, CD-KN2B	rs564398	T	C	Zeggini et al. [34]
	DUSP9	rs5945326	G	A	Voight et al. [27]
	GCC1, PAX4	rs6467136	G	A	Cho et al. [31]
	KIF11	rs6583826	G	A	Sim et al. [24]
	MAEA	rs6815464	C	G	Cho et al. [31]

(continued on next page)

Table 1 – (continued)

GRS	Nearest loci	SNP ID	Risk allele	Other allele	Study used to designate the SNPs into a selected category
	CDKN2B	rs7018475	G	T	Huang et al. [35]
	GLIS3	rs7041847	A	G	Cho et al. [31]
	TMEM45B, BARX2	rs7107217	C	A	Palmer et al. [33]
	C2CD4A, C2CD4B	rs7172432	A	G	Yamauchi et al. [36]
	HMG20A	rs7178572	G	A	Kooner et al. [30]
	C14orf70	rs730570	G	A	Parra et al. [37]
	SLC44A3, F3	rs7542900	C	T	Palmer et al. [33]
	RBM43, RND3	rs7560163	C	G	Palmer et al. [33]
	IRS1	rs7578326	A	G	Voight et al. [27]
	THADA	rs7578597	T	C	Zeggini et al. [25]
	PEXSL	rs7630877	A	G	Sim et al. [24]
	ACHE	rs7636	A	G	Sim et al. [24]
	CDKAL1	rs7754840	C	A	Scott et al. [26]
	TCF7L2	rs7903146	T	C	Saxena et al. [38]
	HNFI1A	rs7957197	T	A	Voight et al. [27]
	TSPAN8, LGR5	rs7961581	C	T	Zeggini et al. [25]
	PRC1	rs8042680	A	C	Voight et al. [27]
	FTO	rs8050136	A	C	Zeggini et al. [34]
	PSMD6	rs831571	C	T	Cho et al. [31]
	JAZF1	rs864745	T	C	Zeggini et al. [25]
	TP53INP1	rs896854	T	C	Voight et al. [27]
	ZFAND3	rs9470794	C	T	Cho et al. [31]
	VEGFA	rs9472138	T	C	Zeggini et al. [25]
	KLF14	rs972283	G	A	Voight et al. [27]

among the Asian and European populations, because of the extreme differential effects of the T2D risk alleles among different population groups that have been noted, for instance, only 10% of the 103 known T2D variants having the same risk allele and direction of effect size in African populations [14–16]. The additive genetic model performs well when the true genetic model is unknown or wrongly indicated [15]. The simple count method also assumes that the SNPs under investigation have similar contribution to T2D risk [15]. The risk alleles of the selected 66 T2D SNPs used in this study were identical to those from which the SNPs were initially discovered as in Table 1. The genotypes of the selected SNPs were assigned a weight of 0, 1 and 2 depending on the number of risk alleles. The 66 SNPs had a proportion of 1.5% of missing values. Due to the limited number of SNPs in our study, which were not genotyped using a SNP array and not strongly linked, the LD methods for imputation were not used. The 66 SNPs were deemed not dense enough compared to SNP arrays that comprise of 0.3 to 5 million SNPs, to form a genotyping scaffold that would enable accurate imputation using LD methods [17]. Imputation accuracy using LD methods has been reported to be adversely affected by SNP density [18]. The series mean imputation method was selected as it had been reported in similar studies with limited SNPs [19,20]. Missing genotype data was imputed with the mean score of the respective SNP. The GRS was determined by summing the scores of the selected SNPs.

Four categories of GRS's from 66 SNPs were calculated to determine the score which was significantly associated with increased risk of T2D in the black South African population as indicated in Table 1. The first GRS was made up of the variants which were previously reported to have statistical

significant associations in the black South African population and it was denoted as GR_{Sn} [12]. Rs8050136, rs831571, rs7542900 and rs1436955 were the variants used to calculate GR_{Sn}. The SNPs reported to be related with beta cell dysfunction among the selected 66 SNPs were selected and a GRS was computed termed GR_{Sbeta}. Rs1436955, rs1387153, rs13266634, rs7754840 and rs7903146 were used to construct GR_{Sbeta} [21,22]. The mechanistic effect of these five SNPs in relation to beta cell dysfunction has been characterised [22]. However, most of the other SNPs have been noted to be in intergenic regions which makes it difficult to characterise the mechanisms underlying their associations with T2D risk [22]. Among the 66 SNPs investigated in this study were variants which had significant associations in GWAS trans-ethnic meta-analysis for T2D, these were used to compute the GRS termed GR_{Strans} [23]. The GR_{Strans} consisted of rs1387153, rs13266634, rs7754840, rs7903146, rs1111875, rs4402960, rs564398, rs8042680, rs864745 variants. The final GRS comprised of all the 66 SNPs and it was depicted as GR_{St}.

2.6. Statistical analysis

The statistical tests were done using SPSS version 21 and Plink version 2.050 [39]. Q-Q plots were used to evaluate the normality of the continuous variables. Chi square tests were used to assess the differences of the categorical variables among the cases and controls. The adherence of the selected variants to the Hardy Weinberg assumptions were evaluated as described previously [12]. The student t tests were used to compare the continuous variables among the cases and controls. Logistic regression was used to assess the association the GRS with type 2 diabetes risk while adjusting for age,

Table 2 – Descriptive statistics of the participants.

Variable	Cases	Controls	p value
Gender n (Male/Females)	39/139	39/139	
Mean age (years)	42.88 ± 4.90	42.51 ± 5.16	0.481
BMI (kg/m ²)	26.10 ± 7.82	22.92 ± 6.02	<0.001
Systolic blood pressure (mmHg)	124.14 ± 18.72	122.02 ± 19.88	0.303
Diastolic blood pressure (mmHg)	84.87 ± 14.16	82.61 ± 13.92	0.132
Fasting glucose (mmol L ⁻¹)	4.92 ± 0.78	4.72 ± 0.66	0.011
HbA1C% (mmol/mol)	5.6 (38) ± 0.5	5.5 (37) ± 0.4	0.022
Urban resident's n (%)	68 (38.2)	39 (21.9)	0.001
2 hr Glucose (mmol L ⁻¹)	9.30 ± 1.40	6.59 ± 0.78	<0.001

HbA1C measure indicated in NGSP% and IFCC mmol/mol; n = number.

Table 3 – Associations of the selected GRSs with risk for type 2 diabetes.

Variable	GRSn			GRSbeta			GRStrans			GRSt		
	OR	95 (CI)	p value	OR	95 (CI)	p value	OR	95 (CI)	p value	OR	95 (CI)	p value
Age (years)	1.01	0.97–1.06	0.597	1.01	0.97–1.06	0.597	1.01	0.97–1.06	0.627	1.01	0.97–1.05	0.671
Gender [†]	1.04	0.61–1.76	0.890	1.07	0.63–1.80	0.856	1.12	0.66–1.90	0.665	1.10	0.65–1.86	0.726
BMI (kg/m ²)	1.07	1.04–1.10	<0.001	1.07	1.03–1.10	<0.001	1.07	1.04–1.10	<0.001	1.07	1.03–1.10	<0.001
GRS [‡]	1.21	1.02–1.43	0.032	0.91	0.78–1.07	0.266	0.84	0.73–0.92	0.020	0.94	0.90–0.99	0.019

Gender[†] = male is reference category; GRS[‡] = category specific genetic risk score respective of the column headings; CI = confidence interval; kg/m² = kilograms per meter squared; OR = odds ratio.**Table 4 – Age and BMI stratified associations of the GRSn with risk for type 2 diabetes.**

Variable	Obese		Non-obese		Less than 50 years		50 years or more	
	OR (95 CI)	p value	OR (95 CI)	p value	OR (95 CI)	p value	OR (95 CI)	p value
Age (years)	1.12 (1.00–1.26)	0.056	1.00 (0.95–1.04)	0.844	1.00 (0.95–1.05)	0.991	1.05 (0.75–1.48)	0.781
Gender [†]	1.18 (0.36–3.88)	0.791	0.91 (0.50–1.66)	0.762	1.02 (0.59–1.77)	0.947	0.59 (0.09–3.84)	0.58
BMI (kg/m ²)	1.01 (0.90–1.13)	0.866	1.07 (1.00–1.14)	0.053	1.07 (1.03–1.10)	<0.001	1.12 (0.96–1.32)	0.159
GRSn	1.18 (0.76–1.82)	0.465	1.21 (1.00–1.46)	0.047	1.23 (1.03–1.47)	0.023	0.89 (0.46–1.73)	0.737

Gender[†] = male is reference category; GRSn = genetic risk score for the significant SNPs associated with type 2 diabetes in this study; CI = confidence interval; kg/m² = kilograms per meter squared; OR = odds ratio.

sex, BMI and urbanisation. The GRS which was the significantly associated with increased T2D risk was selected. The predictive utility of the GRS was assessed by reviewing its improvement on discriminating between the cases and controls using the receiver operated characteristic curves (ROC) and the corresponding areas under the curves (AUC). The ROC of the logistic regression probabilities including conventional T2D risk factors age, sex, BMI and systolic blood pressure were computed alone and in combination with the GRS associated with increased T2D risk. A p-value of 0.05 was used to denote statistical significance.

3. Results

The cases evaluated in the study were significantly more overweight than the controls with a BMI of 26.1 kg/m² compared to the controls' BMI of 22.92 kg/m² (p value < 0.001). A greater proportion of the cases (68%, p value = 0.001) resided in urban areas as opposed to the controls as illustrated in Table 2.

Only GRSn was significantly associated with increased risk of T2D as indicated by an odds ratio (OR) of 1.21 (p value = 0.032). However, the GRStrans and GRSt were significantly less predictive for T2D risk as shown by OR's of 0.84 (p value = 0.02) and 0.94 (p value = 0.019) respectively as illustrated in Table 3.

The interactions of the GRSn with Age and BMI were separately evaluated and the GRSn was noted to have significant interactions of p value = 0.033 and p value < 0.001 with age and BMI respectively. Further explorations indicated the GRSn to be more predictive of risk in participant less than 50 years (OR 1.23; p value = 0.023) and non-obese (OR 1.21; p value = 0.047) compared to those 50 years or more (OR 0.89; p value = 0.737) and the obese (OR 1.18; p value 0.465), respectively as illustrated in Table 4.

The AUC of the ROC was 0.652 ± 0.029 p value < 0.001 for the conventional risk factors and it was modestly increased to 0.665 ± 0.029 p value < 0.001 as a result of the addition of the GRSn as illustrated in Fig. 1. Thereby, indicating the

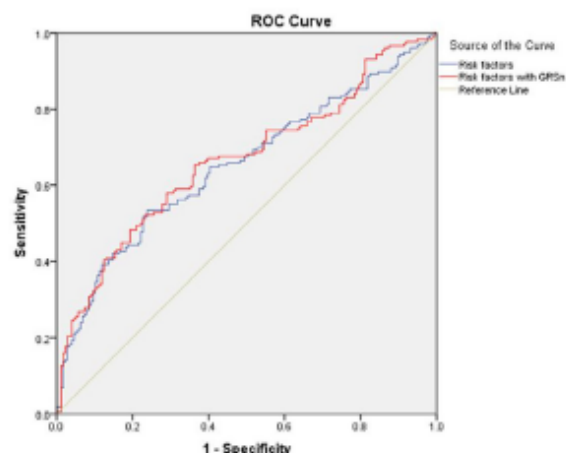


Fig. 1 – Receiver operating curve analysis for selected T2D risk factors. The T2D risk factors (age, urbanisation, BMI, fasting glucose and HbA1C) ability to predict cases alone and in combination with GRSn.

modest improvement in the predictiveness of conventional risk factors with the addition of the GRSn.

4. Discussion

We set out to determine the GRS associated with increased T2D risk in the black South African population of Setswana descent. We found that the GRSn which comprised of variants that had significant associations in this population, to be significantly associated with increased T2D risk as indicated by an OR of 1.21 (1.02–1.43). The GRSn modestly improved the predictive ability of the conventional T2D risk factors as had been reported for other GRS in earlier studies among the Europeans, Asians and African Americans [7,9,40]. The GRS was also noted to be more predictive of risk among the non-obese participants younger than 50 years of age. Thereby, indicating the attractiveness of using the GRS in early identification and prevention strategies.

The GRSn significantly associated with T2D as was expected, since the SNPs which it consisted of, had previously been reported to be associated with T2D in this population group. However, the GRSt consisting of selected 66 the variants in this study was noted to be significantly less predictive [OR 0.95 (0.90–1.00)] against T2D risk, contrary to what was noted in studies among Caucasians where GRSt of more than 60 variants were associated with increased T2D risk as depicted by OR's of 1.13 [41], 1.03 [42] and 1.08 [9]. Similarly, the GRSt was associated with increased risk (OR 1.05) among African Americans [9,40]. The discrepancy between the results of the GRSt in our study and the other ethnicities might be due to the differences in linkage disequilibrium (LD) of the SNPs included in the GRS of the black South Africans with other ethnicities [11]. However, Vassy et al. [9] suggest that the polygenic risk score might be a proxy for the count of the loci which might be associated with T2D among different ethnicities such that the GRSt may be associated with T2D in multiple ethnicities. However, it's still unclear if

the loci are associated with T2D across multiple ethnicities, since no overlap of the chromosomal regions associated with T2D in Caucasians and Africans were noted in linkage association studies, and the causal genes are still yet to be identified [43]. The advent of the customized GWAS genotyping chips in African populations will further help in identifying the population specific variants and variants shared by Africans with other ethnicities associated with T2D as GWAS genotyping chips derived from European ethnicities are less effective in genetic screening of African populations [44,45]. A recent review by Smith et al. [15], suggests that the GRS has to be population specific due to the differences in allele frequencies, LD and ethnicity-specific genetic variation that exist among the varied ethnic groups. Thus, indicating the need to identify population specific GRS for predicting T2D risk among the black South Africans.

It has been suggested that constructing a GRS based on the biological function of variants is more pivotal as compared to using only significant SNPs [9]. The GRS beta which was used in this study was however, associated with protection against T2D and not an increase T2D risk. It is most probable that the approach is valuable as has been documented in other studies, but that the variants associated with beta function in the black South African population might be different [9]. Therefore, it will be pivotal to explore this approach using population specific variants associated with beta function in the future.

There is an understanding that some of the variants associated with T2D are shared among different ethnicities as has been documented in trans-ethnic meta-analysis [46]. Therefore, we selected shared T2D associated variants to develop the GRStrans from one of the largest T2D trans-ethnic meta-analysis for T2D [47]. Nevertheless, the GRStrans was significantly less predictive against T2D risk in our study population. In view that the GRStrans, did not comprise of variants discovered in black South Africans, differences in LD as noted for the GRSt might have led to it being less predictive [11,15,48].

The GRSn had significant interactions with Age and BMI in separately of p values equal to 0.033 and <0.001 respectively. These interactions were further explored in Age and BMI stratified associations of the GRSn with T2D risk. The GRSn was noted to be more predictive to risk among less than 50 years old and non-obese participants compared to the obese and 50 years or more participants as indicated in Table 4. Though this finding among this population group maybe first being indicated in our study, it has been reported among the Asian and European population groups [8,49].

In view that T2D is heritable, it is expected to depict the genetic risk early in life before the development of other risk factors such as obesity have been noted. The GRSn is a proxy of the T2D genetic risk and hence its improved predictiveness among the less 50-year-old participants was expected, and other GRS proxies by de Miguel-Yanes et al. also reported similar results [8]. Evidence from a longitudinal study in which a GRS was noted to be more predictive of T2D risk in young adults aged 18–30 years for a period of for 25 years provides strong support of the GRS approach being more predictive in risk in people less than 50 years as depicted in our study [50].

With regards to the GRS interactions with obesity, Perry et al. reported a per risk allele OR of 1.13 (p value = 2.2×10^{-14}) among the lean T2D cases as opposed to an OR of 1.06

(p value = 2.2×10^{-16}) in the obese among the Caucasians indicating the improved predictiveness of the GRS among lean T2D patients [51]. Similarly this phenomena has recently been reported by Kong et al. in a Han Chinese population [49]. The high heterogeneity of T2D is postulated to be responsible for the differences in genetic risks among the obese and non-obese cases. Epidemiological studies indicate that the non-obese patients develop rapid β cell failure in the condition of insulin sensitivity thereby requiring early insulin treatment compared to the obese [52]. This heterogeneity might be as result of different factors which include genetics being involved in the T2D etiology among obese and non-obese patients [53]. The majority of the known T2D variants are postulated to be related to beta cell function, thereby explaining the improved predictiveness of the GRS in the lean patients who have been noted to be more susceptible to β cell failure in epidemiological studies. However, since the function of the variants of the SNPs in our GRS which was more predictive among the non-obese patients in our study is not known as they reside in intergenic regions there might be plausibly related to β cell function.

5. Conclusion

In view of our findings, the GRS approach is of less clinical use in the black South African population due to its modest improvement in predicting T2D risk. Since T2D, has a polygenic etiology, the GRS approach is being further limited by the few variants significantly associated with T2D in the black South African context. Thus, direct consumer genetic testing is of less clinical use in the black South African population as has been documented in European and Asian ethnicities [7,10]. The advent of the Afrocentric customized genotyping array gives hope for the possibility of identifying more population sensitive variants which might improve the utility of the GRS in the future. The short LD blocks among African populations, ensure an increased chance of identify the causal variants in future GWAS in Africa. These causal variants combined into a GRS might be more predictive of T2D risk [11]. The pathway based analysis of these population variants and inclusion of rare variants in the GRS might in the future yield a more clinically useful GRS for early identification of high risk T2D patients.

Conflict of interest statement

This research was supported by a grant National Research Foundation of South Africa (87970). Tinashe Chikwore is supported by the National Research Foundation of South Africa (89117). We would like to acknowledge Prof A Kruger and the PURE research team, funders and participants. All the authors have no conflicts of interest to declare.

REFERENCES

- [1] Bradshaw D, Norman R, Pieterse D, Levitt NS. *S Afr Med J* = Suid-Afrikaanse tydskrif vir geneeskunde. Suid-Afrikaanse tydskrif vir geneeskunde. 2007;97:700–6.
- [2] Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract* 2011;94:311–21.
- [3] Groop L, Pocot F. Genetics of diabetes—are we missing the genes or the disease? *Mol Cell Endocrinol* 2014;382:726–39.
- [4] Saudek CD, Herman WH, Sacks DB, Bergenstal RM, Edelman D, Davidson MB. A new look at screening and diagnosing diabetes mellitus. *J Clin Endoc Metab* 2008;93:2447–53.
- [5] Florez JC. Chapter 82 – diabetes. In: Ginsburg GS, Willard HF, editors. *Genomic and personalized medicine*. Academic Press; 2013.
- [6] Pyke DA. Development of diabetes in identical twins. *Adv Exp Med Biol* 1988;246:255–8.
- [7] Wang X, Strizich G, Hu Y, Wang T, Kaplan RC, Qi Q. Genetic markers of type 2 diabetes: progress in genome-wide association studies and clinical application for risk prediction. *J Diabetes* 2016;8:24–35.
- [8] de Miguel-Yanes JM, Shrader P, Pencina MJ, Fox CS, Manning AK, Grant RW, et al. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes Care* 2011;34:121–5.
- [9] Vassy JL, Hivert MF, Porneala B, Dauriz M, Florez JC, Dupuis J, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes* 2014;63:2172–82.
- [10] Bao W, Hu FB, Rong S, Rong Y, Bowers K, Schisteman EF, et al. Predicting risk of type 2 diabetes mellitus with genetic risk models on the basis of established genome-wide association markers: a systematic review. *Am J Epidemiol* 2013;178:1197–207.
- [11] Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet* 2010;11:149–60.
- [12] Chikwore T, Conradie KR, Towers GW, van Zyl T. Common variants associated with type 2 diabetes in a Black South African population of setswana descent: African populations diverge. *OMICS* 2015;19:617–26.
- [13] Müller B, Wilcke A, Boulesteix AL, Brauer J, Passarge E, Boltze J, et al. Improved prediction of complex diseases by common genetic markers: state of the art and further perspectives. *Hum Genet* 2016;135:259–72.
- [14] Chen R, Corona E, Sikora M, Dudley JT, Morgan AA, Moreno-Estrada A, et al. Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet* 2012;8:e1002621.
- [15] Smith JA, Ware EB, Middha P, Beacher L, Kardis SL. Current applications of genetic risk scores to cardiovascular outcomes and subclinical phenotypes. *Curr Epidemiol Rep* 2015;2:180–90.
- [16] Adeyemo AA, Tekola-Ayele F, Doumatey AP, Bentley AR, Chen G, Huang H, et al. Evaluation of genome wide association study associated type 2 diabetes susceptibility loci in Sub Saharan Africans. *Front Genet* 2015;6:335.
- [17] Nelson SC, Doheny KF, Pugh EW, Romm JM, Ling H, Laurie CA, et al. Imputation-based genomic coverage assessments of current human genotyping arrays. *G3 (Bethesda)* 2013;3:1795–807.
- [18] Morris AP. Fine mapping of type 2 diabetes susceptibility loci. *Curr Diab Rep* 2014;14:549.
- [19] Fontaine-Bisson B, Renstrom F, Rolandsson O, Magic Payne F, Hallmans G, et al. Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population. *Diabetologia* 2010;53:2155–62.
- [20] Poveda A, Koivula RW, Ahmad S, Barroso I, Hallmans G, Johansson I, et al. Innate biology versus lifestyle behaviour in the aetiology of obesity and type 2 diabetes: the GLACIER study. *Diabetologia* 2016;59:462–71.

- [21] Qi Q, Hu FB. Genetics of type 2 diabetes in European populations. *J Diabetes* 2012;4:203–12.
- [22] Thomsen SK, Gloyn AL. The pancreatic beta cell: recent insights from human genetics. *Trends Endocrinol Metab* 2014;25:425–34.
- [23] Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012;44:981–90.
- [24] Sim X, Ong RT, Suo C, Tay WT, Liu J, Ng DP, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet* 2011;7:e1001363.
- [25] Zeggini E, Scott LJ, Saxena R, Voight BF. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40:638–45.
- [26] Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science* 2007;316:1341–5.
- [27] Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010;42:579–89.
- [28] Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proença C, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* 2009;41:1110–5.
- [29] Shu XO, Long J, Cai Q, Qi L, Xiang YB, Cho YS, et al. Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet* 2010;6:e1001127.
- [30] Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 2011;43:984–9.
- [31] Cho YS, Chen CH, Hu C, Long J, Ong RT, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 2011;44:67–72.
- [32] Tsai J, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet* 2010;6:e1000847.
- [33] Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, An SS, et al. A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS ONE* 2012;7:e29202.
- [34] Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;316:1336–41.
- [35] Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur J Hum Genet* 2012;20:801–5.
- [36] Yamauchi T, Hara K, Maeda S, Yasuda K, Takahashi A, Horikoshi M, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A–C2CD4B. *Nat Genet* 2010;42:864–8.
- [37] Parra EJ, Below JE, Krithika S, Valladares A, Barta JL, Cox NJ, et al. 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* 2011;54:2038–46.
- [38] Diabetes Genetics Initiative of Broad Institute of H, Mit LU, Novartis Institutes of Biomedical R, Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–6.
- [39] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- [40] Imamura M, Shigemizu D, Tsunoda T, Iwata M, Maegawa H, Watada H, et al. Assessing the clinical utility of a genetic risk score constructed using 49 susceptibility alleles for type 2 diabetes in a Japanese population. *J Clin Endocrinol Metab* 2013;98:E1667–73.
- [41] Walford GA, Porneala BC, Dauriz M, Vassy JL, Cheng S, Rhee EP, et al. Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. *Diabetes Care* 2014;37:2508–14.
- [42] Vaxillaire M, Yengo L, Lobbens S, Rocheleau G, Eury E, Lantieri O, et al. Type 2 diabetes-related genetic risk scores associated with variations in fasting plasma glucose and development of impaired glucose homeostasis in the prospective DESIR study. *Diabetologia* 2014;57:1601–10.
- [43] Ng MCY. Genetics of type 2 diabetes in African Americans. *Curr Diabetes Rep* 2015;15:74.
- [44] Dandara C, Huzair F, Borda-Rodriguez A, Chirikure S, Okpechi I, Warnich L, et al. H3Africa and the African life sciences ecosystem: building sustainable innovation. *OMICS* 2014;18:733–9.
- [45] May A, Hazelhurst S, Li Y, Norris SA, Govind N, Tikly M, et al. Genetic diversity in black South Africans from Soweto. *BMC Genomics* 2013;14:644.
- [46] Waters KM, Stram DO, Hassanein MT, Le Marchand L, Wilkens LR, Maskarinec G, et al. Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups. *PLoS Genet* 2010;6:10.
- [47] Replication DG, Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes C, South Asian Type 2 Diabetes C, Mexican American Type 2 Diabetes C, Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples C, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014;46:234–44.
- [48] Mhandire K, Pharo G, Kandawasvika GQ, Duri K, Swart M, Stray-Pedersen B, et al. How does mother-to-child transmission of HIV differ among African populations? Lessons from MBL2 genetic variation in Zimbabweans. *OMICS* 2014;18:454–60.
- [49] Kong X, Xing X, Hong J, Zhang X, Yang W. Genetic variants associated with lean and obese type 2 diabetes in a Han Chinese population: a case-control study. *Medicine (Baltimore)* 2016;95:e3841.
- [50] Vassy JL, Durant NH, Kabagambe EK, Carnethon MR, Rasmussen-Torvik LJ, Fornage M, et al. A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia* 2012;55:2604–12.
- [51] Perry JR, Voight BF, Yengo L, Amin N, Dupuis J, Ganser M, et al. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet* 2012;8:e1002741.
- [52] George AM, Jacob AG, Fogelfeld L. Lean diabetes mellitus: an emerging entity in the era of obesity. *World J Diabetes* 2015;6:613–20.
- [53] Timpson NJ, Lindgren CM, Weedon MN, Randall J, Ouwehand WH, Strachan DP, et al. Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes* 2009;58:505–10.

APPENDIX C

Acceptance letter for publication: Nutrient patterns associated with fasting glucose and glycated hemoglobin in a black South African population.



Tinashe Chikowore <tinashedoc@gmail.com>

[Nutrients] Manuscript ID: nutrients-155427 - Accepted for Publication

Annie Wang <annie.wang@mdpi.com>

Mon, Dec 19, 2016 at 2:52 AM

Reply-To: annie.wang@mdpi.com

To: Tinashe Chikowore <tinashedoc@gmail.com>

Cc: Pedro Pisa <ppisa@wrhi.ac.za>, Tertia Van Zyl <tertia.vanzyl@nwu.ac.za>, Edith Feskens <edith.feskens@wur.nl>, Edelweiss Wentzel-Viljoen <edelweiss-wentzel-viljoen@nwu.ac.za>, Karin Conradie <karin.conradie@nwu.ac.za>, Nutrients Editorial Office <nutrients@mdpi.com>

Dear Dr, Chikowore,

We are pleased to inform you that the following paper has been officially accepted for publication:

Manuscript ID: nutrients-155427

Type of manuscript: Article

Title: Nutrient patterns associated with fasting and glycated hemoglobin levels among urban black South African women

Authors: Tinashe Chikowore *, Pedro Pisa, Tertia Van Zyl, Edith Feskens, Edelweiss Wentzel-Viljoen, Karin Conradie

Received: 27 September 2016

E-mails: tinashedoc@gmail.com, ppisa@wrhi.ac.za, tertia.vanzyl@nwu.ac.za, edith.feskens@wur.nl, edelweiss-wentzel-viljoen@nwu.ac.za, karin.conradie@nwu.ac.za

Nutrition and Diet Factors in Type 2 Diabetes

http://www.mdpi.com/journal/nutrients/special_issues/type_2_diabeteshttp://susy.mdpi.com/user/manuscripts/review_info/a0b72eaa49a587399accbd3388964d41

We will now make the final preparations for publication, then return it to you for your approval.

Kind regards,
Annie

Ms. Annie Wang
Assistant Editor

E-Mail: annie.wang@mdpi.comNutrients (<http://www.mdpi.com/journal/Nutrients>)

—

Ms. Annie Wang

MDPI Branch Office, Beijing

8th Floor, Aerospace Cooperation Building, No.99 Zhongguancun East Road, Haidian District, Beijing, 100190, China

Tel./Fax: +86 10 5947 3709

MDPI AG

Postfach, CH-4020 Basel, Switzerland

Office: St. Alban-Anlage 66, 4052 Basel

Tel. +41 61 683 77 34 (office)

Fax +41 61 302 89 18 (office)

Nutrients Editorial Office