GENETIC AND NON-GENETIC STUDIES OF TYPE 2 DIABETES IN THREE SUSCEPTIBLE ASIAN POPULATIONS: MALAY, CHINESE AND INDIAN

NORAIDATULAKMA ABDULLAH
MScCH (Epidemiology and Medical Statistics),
B.Sc.(Hons) (Genetics)

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
SCHOOL OF BIOMEDICAL SCIENCES AND PHARMACY
THE UNIVERSITY OF NEWCASTLE
AUGUST 2017

ABSTRACT

Both genetic and non-genetic factors have been reported to contribute to the pathogenesis of type 2 diabetes. Although numerous epidemiological studies have been conducted in various populations, the Malaysian society remains relatively understudied to date, despite having a relatively high prevalence of type 2 diabetes among Asian countries. Within Malaysia, the type 2 diabetes prevalence also differs between major ethnic groups, being highest in Indian, intermediate in Malays and lowest in Chinese. To better understand the relative contributions of genetic and non-genetic risk factors to type 2 diabetes in Malaysia, this study conducted epidemiological studies of type 2 diabetes in Malaysian participants of Malay, Chinese and Indian ethnicity from The Malaysian Cohort project.

Samples from 1,604 Malays, 1,654 Chinese and 1,728 Indians were included in genetic analyses, which used genotyped data obtained from the Metabochip array. A total of 62 individual candidate single nucleotide polymorphisms (SNPs) previously associated with type 2 diabetes were assessed, individually and in the form of a genetic risk score aggregating information across all polymorphisms. Utilising the same samples, the effects of environmental (non-genetic, or lifestyle) risk factors were also assessed. Finally, we assessed the evidence for effect modification of environmental effects by genetic alleles (gene by environment).

After Bonferroni correction for multiple testing, seven (7) individual SNPs showed association with type 2 diabetes in analyses of the combined Malaysian sample, adjusted for ancestry. An additional 10 SNPs showed nominal association ($p<0.05$ before adjustment for multiplicity). The genetic risk score showed strong association with type 2 diabetes in the individual ancestral groups (p-values ranging from 4.71×10^{-6} to 1.35×10^{-8}), and the combined group ($p=2.2 \times 10^{-16}$). However, the genetic risk score explained only 1.0 to 1.7% of total risk variance. In contrast, four non-genetic risk factors, age, gender, waist-to-hip ratio and physical inactivity, accounted for about 20% of total type 2 diabetes risk variation in the Malaysian samples. The effect of increasing waist-to-hip ratio was higher in Chinese than Indian or Malay participants, suggesting anthropometric risk differences between groups.

Incorporating the genetic risk score into statistical models including the environmental factors only explained an additional 1 to 2% of risk variation in each group. We found some evidence for gene by environment effect modification, with the genetic risk score showing a gradient of decreasing effect sizes across increasing strata of body mass index. While formal tests of interaction were non-significant, this is consistent with previous evidence and suggests genetic risk factors may have a larger contribution to disease pathogenesis in leaner type 2 diabetes cases. Taken together, these studies suggest that environmental, rather than genetic risk factors are the major contributors to the epidemic of type 2 diabetes in Malaysia.

Our findings have some public health significance in relation to mitigating type 2 diabetes risk in Malaysia. First, these findings may inform targeted interventions focussing on abdominal obesity in the

Malaysian population, especially in Chinese Malaysians. Second, these results suggest a need for the development of ethnicity-specific anthropometric cut-points, to accurately assess associations across ancestral groups with different body fat distributions. Third, these findings suggest a relatively greater contribution of genetic factors to disease among genetically predisposed lean individuals, which may have implications for personalised medicine. Future studies in larger samples could similarly investigate these findings, to further clarify the respective roles of genetic and environmental risk factors to disease, and inform personalised interventions.

OVERVIEW

Identifying genetic, environmental risk factors and potential interactions between them may provide insights into factors contributing to the rapid increase of type 2 diabetes (T2D) prevalence in Malaysian populations, and in turn to identify targeted interventions that may reduce the burden of T2D. Genetic studies in diverse populations are also vital to ascertain the factors contributing to the disparity of T2D population prevalence among Malaysian ethnic groups, considering that they are sharing a similar environment. Although numerous studies have been performed to identify genetic and environmental risk factors in T2D, no large-scale studies have been performed in a Malaysian population, despite its highest comparative prevalence of T2D among Asian countries.

This thesis is structured as "Thesis by Publication". Chapter 1 provides an introduction and background to the study, defines the problem, objectives and methodology. Chapters 2 to Chapter 5 are a compilation of publications representing as outlined below.

Chapter 2 is a review article serving as literature review to identify and compile previous work on genetic and environmental risk factors for T2D in diverse populations. This review article included the compilation and description of 118 genetic risk variants found to be significantly associated with T2D in various populations. This review also highlighted the importance and value of genetic studies in participants from multi-ethnic background.

Chapter 3 describes a genetic study of T2D in the Malaysian population. This chapter addresses the first objective of this study. At the time of writing of this thesis, this is the largest scale of genetic study in T2D performed in a Malaysian population. This study confirmed the involvement of seven individual T2D genetic variants in the Malaysian population and additional ten individual genetic variants that reach nominal significance. However, a genetic risk score aggregating 62 SNPs explained less than 2% of total T2D variation in the Malaysian population, demonstrating a substantial contribution by additional risk factors.

Chapter 4 investigates the contribution of environmental risk factors to T2D risk in the Malaysian population. This chapter addresses the second objective of this study. The risk factors assessed in this study included demographic, lifestyle and anthropometric measurements. The combination of four non-genetic risk factors: age, gender, waist-to-hip ratio (WHR) and physical inactivity, accounted for about 20% of T2D risk in the combined Malaysian sample. This indicated that major contributors to the increasing T2D prevalence in Malaysia are determinants of obesity such as diet and physical inactivity, together with the ageing population. The predictive accuracy (Area Under the Receiver Operator Characteristic Curve) of the four risk factors were ranging from 0.75 to 0.83; being lowest in Malays and highest in Chinese ancestry. The disproportion of AUC across the ancestry groups was due to population differential effects of waist-to-hip ratio (WHR), which may reflect ancestral differences in body fat percentage.

Chapter 5 addresses the third objective of this study, assessing gene-environment interaction in T2D. Interaction analyses were performed for both individual SNPs and genetic risk score (GRS), on both multiplicative and additive scales. Although this study found null interaction in both individual SNPs and GRS, some evidence of genetic effect gradient across BMI strata with inversed relationship was observed. This proposed that lean T2D cases may have a higher genetic predisposition to T2D than overweight or obese cases. Significant improvement in T2D risk explained and predictive risk due to the GRS was observed although only minimal increment about 1-2% in pseudo R^2 and 1-3% in AUC respectively were observed. Such a small increment reflects that common genetic variants with small effects are involved in the pathogenesis of T2D.

Chapter 6 provides an overall discussion of this body of work. This chapter discusses the strengths, limitations and the contribution of this research to the field. It also suggests future directions of epidemiological research for T2D in a multi-ethnic country such as Malaysia.

DECLARATION

I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision.

The thesis contains published scholarly work of which I am a co-author. For each such work a written statement, endorsed by the other authors, attesting to my contribution to the joint work has been included.

The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

NORAIDATULAKMA ABDULLAH

23 July 2017

STATEMENT OF ORIGINALITY

*This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis, being made available worldwide when deposited in the University's Digital Repository**, subject to the provisions of the Copyright Act 1968.*

***Unless an embargo has been approved for determined period.*

NORAIDATULAKMA ABDULLAH

23 July 2017

STATEMENT OF COLLABORATION

I hereby certify that the work embodied in this thesis has been done in collaboration with other researchers, or carried out in other institutions. I have included as part of the thesis a statement clearly outlining the extent of collaboration, with whom and under what auspices.

NORAIDATULAKMA ABDULLAH

23 July 2017

STATEMENT OF AUTHORSHIP

I hereby certify that this thesis is in the form of a series of published papers of which I am a joint author. I have included as part of the thesis a written statement from each co-author, endorsed by the Faculty Assistant Dean (Research Training), attesting to the joint publications.

NORAIDATULAKMA ABDULLAH

23 July 2017

ACKNOWLEDGEMENT

Completing this Doctoral research has been one of the biggest academic challenges I have faced in my life. It would not have been possible without the great help, support, guidance and assistance of numerous people throughout the project. First and foremost I would like to thank Almighty God, the compassionate, the Almighty Merciful, who kindly helped me throughout this PhD journey and to complete my thesis.

First of all, I would like to express my deepest gratitude and appreciation to my Principal Supervisor Dr Elizabeth Holliday, who has accompanied me throughout this challenging journey. Her guidance, experience, and knowledge were a great aid during my research work as well as while writing the manuscripts and this thesis. She acted as my mentor for the last few years despite her many other academic and professional commitments. Her patience, kindness and support helped me to get through my PhD journey, despite adverse circumstances.

I would like to express my gratitude to my co-supervisors, Laureate Professor Rodney Scott, Professor John Attia and Dr Christopher Oldmeadow for providing support, guidance and professional advice throughout the duration of the research. I am grateful for their assistance. I would like to express my gratitude and appreciation to my employer, Professor Datuk A. Rahman A. Jamal, the director of UKM Medical Molecular Biology Institute (UMBI) at the Universiti Kebangsaan Malaysia (UKM) and project leader for The Malaysian Cohort Project, for allowing me to use data from the Malaysian Cohort Project for my PhD. My thanks go to all UMBI and The Malaysian Cohort staff for providing the Type 2 diabetes GWAS data especially Dr Azian Abdul Murad, Dr. Ezanee Azlina Mohammad Hanif, Mrs Nazihah Jalal, Mrs Norliza Ismail and Mrs Afzan Effiza Abdul Patah.

I would like to mention my appreciation to the members of the Centre of Information Based Medicine research group, especially Miss Trish Collison, Dr Vicki Maltby, Dr Michelle Wong-Brown, Dr Kelly Kiejda, Miss Tiffany Evans, Miss Chloe Warren, Miss Andrea Mathe, and Miss Brianna Morten, for their support throughout and for making my stay in Newcastle enjoyable.

Last but not least, I would like to express the appreciation from all of my heart to my family and especially my aunt, Dr Norhayati Soin, for their incredible support, love, and help with every step of the process. I always knew they prayed for me, they love me, and will do everything they can to help me succeed. I am speechless when I think about everything my family has done to make all this happen.
I would like to say thank you to all those I am unable to name who participated in this success and without whom I would have been unable to complete this work.

Finally, I would like to dedicate this thesis to my late parents, Mr Abdullah Mohd Zain and Mrs Khatijah Embong, and my late husband, Ir. Sufyan Samsuddin, who always supported me to pursue my PhD. My special dedication goes to my two beautiful daughters, Sarah and Salma, for giving me strength and making the journey bearable.

LIST OF PUBLICATIONS INCLUDED AS PART OF THIS THESIS

1. **Abdullah N**, Attia J, Oldmeadow C, Scott RJ, Holliday EG. The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings. *International Journal of Endocrinology*. 2014; 2014:593982. doi: 10.1155/2014/593982. Epub 2014 Mar 13. (PMID:24744783).
2. **Abdullah N**, Abdul Murad N. A, Attia J, Oldmeadow C, Mohd Haniff E. A, Syafruddin S. E, Abd Jalal N, Ismail N, Ishak M, Jamal R, Scott R.J, Holliday E. G, Characterising the genetic risk for type 2 diabetes in a Malaysian multi-ethnic cohort. *Diabetes Medicine*. 2015 Oct; 32 (10):1377-84. doi: 10.1111/dme.12735. Epub 2015 Mar 24 (PMID: 25711284).
3. **Abdullah N**, Abdul Murad N. A, Attia J, Oldmeadow C, Kamaruddin M.A., Abd. Jalal N., Ismail N., Jamal R, Scott R.J, Holliday E. G. Quantifying the Roles of Classical Risk Factors in Type 2 Diabetes using a Multi-ethnic Malaysian Cohort. *Submitted*.
4. **Abdullah, N**., Abd Murad, NA., Mohamad Haniff, EA., Attia, J., Oldmeadow, C., Syafrudin, SE., Kamaruddin, MA., Ismail, N., Jalal, N., Ishak, M., Jamal, R., Holliday, E. Predicting Type 2 Diabetes using genetic and environmental risk factors in a multi-ethnic Malaysian Cohort. Public Health 149 (2017)31-38. April 2017. http://dx.doi.org/10.1016/j.puhe.2017.04.003 (PMID:28528225).

OTHER PUBLICATIONS AND PRESENTATIONS ARISING FROM THIS THESIS

1. Cohort Profile: The Malaysian Cohort (TMC) project: a prospective study of non-communicable diseases in a multi-ethnic population. Jamal R, Syed Zakaria SZ, Kamaruddin MA, Abd Jalal N, Ismail N, Mohd Kamil N, **Abdullah N**, Baharudin N, Hussin NH, Othman H, Mahadi NM; Malaysian Cohort Study Group. Int J Epidemiol. 2015 Apr;44(2):423-31. doi: 10.1093/ije/dyu089. Epub 2014 Apr 11. (PMID: 24729425).
2. Poster Presented at 7th Australian Health & Medical Research Congress (AHMRC) 2014, 16th - 19th November 2014, Melbourne, Australia.
3. Oral Presented at the Australian Society for Medical Research (ASMR) Satellite Scientific Meeting 2015, 21st March 2015, Newcastle, Australia.
4. Poster Presented at 75th American Diabetes Association (ADA) Scientific Sessions 2015, 5th -9th June 2015, Boston, USA.

LIST OF FIGURES

GLOSSARY OF ABBREVIATIONS AND ACRONYMS

Abbreviated Term	Expanded Term
T2D	Type 2 Diabetes
mmol/L	Millimoles Per Litre
mg/L	Milligrams Per Litre
mg/dL	Milligrams Per Decilitre
OGTT	Oral Glucose Tolerance Test
HbA1c	Haemoglobin A1c
CGAS	Candidate-gene Association Studies
GWAS	Genome wide Association Study
SNP	Single Nucleotide Polymorphism
NHMS	National Health Morbidity Survey
MCP	The Malaysian Cohort Project
NCBI	National Centre of Biotechnology Information
NIH	National Institute of Health
NHGRI	National Human Genome Research Institute
UKM	Universiti Kebangsaan Malaysia
MAF	Minor Allele Frequency
HWE	Hardy-Weinberg Equilibrium
PCA	Principal Component Analysis
SGVP	Singaporean Genome Variation Project
GRS	Genetic Risk Score
BMI	Body Mass Index
WC	Waist Circumference
WHR	Waist-to-Hip Ratio
IPAQ-M	International Physical Activity Questionnaire
MICE	Multiple Imputations by Chained Equations
MAR	Missing At Random
AUROC	Area Under the Receiver Operating Characteristic
RERI	Relative Excess Risk Due to Interaction
EPIC	European Prospective Investigation Into Cancer and Nutrition
SIGMA	Slim Initiative in Genomic Medicine for the Americas
JBASE	Joint Bayesian Analysis
UKPDS	UK Prospective Diabetes Study
DCCT	Diabetes Control Complications Trials
miRNAs	MicroRNAs
piRNAs	PIWI-Interacting RNAs
snoRNAs	Small Nuclear RNAs
lincRNAs	Long intergenic non-coding RNAs
lncRNAs	Long non-coding RNAs

TABLE OF CONTENT

CHAPTER 1: INTRODUCTION

1.1 Background: Type 2 Diabetes

Type 2 diabetes (T2D) is a consequence of the ineffective utilization of insulin due to insulin resistance and/or reduced insulin availability in the body. Insulin is produced by beta cells within the islets in the pancreas and facilitates the transfer of sugar from the bloodstream into cells. The efficiency of glucose clearance from the bloodstream is thus dependent on insulin concentrations in the blood. As a result of impairments in both insulin release from beta cells and insulin sensitivity, people with T2D have elevated blood sugar. T2D is diagnosed by several diagnostic tests: having a fasting blood glucose concentration greater than or equal to 7 mmol/L or 126 mg/dL, or non-fasting glucose greater than or equal to 11.1 mmol/L or 200 mg/dL, or glucose greater than or equal to 11.1 mmol/L or 200 mg/dL in a 2-hour glucose tolerance test (OGTT), or a haemoglobin A1C (HbA1c) test greater than or equal to 6.5%.

T2D is a chronic disease of generally long duration and slow progression which can remain unnoticed and undiagnosed for years. Once noticed, complications may already be severe, as a result of sustained elevations in blood glucose. Common complications of T2D include nephropathy (kidney failure), retinopathy (loss of vision), cardiovascular disease, stroke and damage to nerves and blood vessels. Such damage can decrease blood circulation, leading to amputation of the toes, feet or legs. It also can increase susceptibility to other diseases and thus the development of comorbidities, loss of mobility with aging, and the risk of depression [1, 2].

Currently, it is estimated that 415 million people are living with diabetes worldwide, with the vast majority of cases being T2D [3]. Globally, the prevalence of T2D has risen from 4.7% in 1980 to 8.5% in 2014 [4]. In the Asian context, T2D prevalence has been rising particularly rapidly in recent years, with about 60% of T2D patients worldwide now residing in Asia. Asian patients with T2D also appear more prone to developing complications than Europeans with regard to renal failure, cardiovascular disease and stroke [5]. Due to the prevalence of T2D and the severity of its complications, the disease carries a substantial economic burden. In 2015, approximately 12% of the Asia region's total health budget was spent on T2D-related care including treatment of T2D and related complications [6]. In most Asian countries, the economic burden of T2D is enormous and unmatched by budget allocations for health care. This is due to the combined effects of increased health service use, lost productivity and long-term carer support required to manage T2D-related complications [7]. Population growth, in parallel with increasing urbanisation and lifestyle changes, is expected to further increase the burden of disease and future health expenditure associated with T2D in Asian countries in future years [8].

Compared to persons from European populations, who have a comparatively lower prevalence of T2D, Asians often develop T2D at lower BMI and younger age. The reasons for this are not entirely clear. The aetiology of T2D is complex and includes influences from both environmental and genetic risk factors [9]. In Asia, the rising prevalence of T2D has been associated with factors including reduced

physical activity, rapid socio-economic development and urbanization, dietary changes, and smoking [7]. Worldwide, the prevalence of T2D has increased in parallel with increases in obesity prevalence, with obesity being a known contributor to increased insulin resistance and T2D development and progression. However, not all individuals with T2D are overweight or obese, and conversely, many obese people do not develop T2D [10]. Thus, obesity is neither necessary nor sufficient for T2D development, highlighting the role of other risk factors.

In addition to lifestyle risk factors such as obesity and physical inactivity, T2D risk also has a substantial genetic component. Estimates of the heritability of T2D are on the order of 30-70%, based on family and twin studies, [11]. This means that about 30-70% of the observed variation in T2D risk can be attributed to the effects of genetic variation. The frequency of genetic risk variants, and their influence in the context of particular lifestyle factors, may differ between ancestral groups. In this way, genetic variation could contribute to population differences in the prevalence of heritable traits such as T2D. For example, according to the National Diabetes Statistics Report, the prevalence of T2D among population groups in the United States varies from 7.6%-15.9% [12]. The lowest prevalence occurs in non-Hispanic whites (7.6%) and the highest prevalence occurs in American Indians/Alaskan Natives (15.9%). Among American Indians, the Pima Indians of Arizona have highest prevalence, with approximately 50% of adults above 35 years having T2D [12]. The disparity of T2D prevalence between ethnic groups sharing a similar environment suggests that genetic factors (which differ between ethnic groups) may contribute to these differences. The increased T2D prevalence in Asian countries and the clear disparity in prevalence between Asian and non-Asian population groups provide a justification for detailed genetic studies of T2D in Asian populations. Moreover, because the prevalence of T2D also differs between Asian populations, detailed study of diverse Asian groups is warranted, to better understand differences in risk.

1.1.1 Genome-wide association studies of T2D

In spite of its high heritability, identifying genetic factors influencing T2D risk has been challenging. Nevertheless, substantial progress has been made in recent years. In the late 20th century, the identification of genetic risk variants was primarily limited to rare, monogenic forms of T2D [13]. However, in about the last decade, as human genome sequence information became available, together with advanced genomic technologies for high throughout genotyping, genome-wide association studies (GWAS) have enabled a new era of discovery. Results from genome-wide association studies (GWAS) have emphasised the complexity and polygenic nature of T2D, and also provided insights into the similarities and differences in its genetic architecture between population groups [14]. A variety of T2D GWAS have been conducted in about the last ten years, collectively identifying and validating numerous genetic associations for the disease. Key studies and their findings will be reviewed here.

The first study conducted by Sladek and colleagues [15] in 2007 utilized 600 cases and 600 controls of European ancestry and reported T2D-associated variants in three susceptibility genes: *HHEX/IDE*, *SLC30A8* and *TCF7L2*. The genes *HHEX/IDE* and *TCF7L2* are associated with β- cell function [16] while *SLC30A8* encodes a zinc transporter significantly expressed in pancreatic islets [17]. Subsequent GWAS identified another four T2D variants in 2007 [18-22]. These variants were within *CDKN2A/CDKN2B* and *CDKAL1*, which are associated with β- cell development [23, 24], *IGF2BP2* variants associated with β- cell dysfunction [23] and *FTO*, which has indirect effects on T2D via obesity [25, 26]. The first wave of T2D GWAS culminated in the first GWAS meta-analysis of T2D [27] and the first non-European GWAS conducted in Japanese populations [28]. Meta-analysis is a cost- and time-effective way to increase sample size by combining data from multiple individual GWAS, often being performed by dedicated consortia. The larger effective sample size increases the statistical power of the study and the ability to identify genetic risk variants of small effect.

A key finding among initial GWAS was the discovery of *KCNQ1* variants in an Asian population, highlighting the utility of extending T2D GWAS to non-European populations. This variant was not detected in previous European GWAS due to a significantly lower frequency of the risk allele in Europeans compared to East Asian (5% versus 40%), showing that population differences in risk allele frequency can dramatically affect statistical power for variant discovery [29].

The second wave of T2D GWAS began in 2010, with additional GWAS in non-European population being conducted. In 2010, three T2D GWAS were conducted in East Asian populations (Japanese and Chinese populations) and one in an African American population [30-33]. The following year, (2011), five T2D GWAS were conducted in non-European populations. One meta-analysis of South East Asian populations included samples from the Malay, Chinese and Indian population, while other studies were performed in South Asian, Mexican American and Hispanic groups [14, 34-36]. These GWAS conducted in diverse population groups discovered 21 novel loci in spite of relatively modest sample sizes, again underscoring the utility of assessing different populations. A large meta-analysis in 2012 identified a T2D risk variant in the *MAEA* gene. This was found to be unique to African and East Asian populations, with the variant being monomorphic in Europeans and South Asians [37, 38]. A similar finding was reported for a *SCGG* variant, found to be unique to the Indian Punjabi Sikh, and being monomorphic in both European and African populations [39].

A comprehensive review of T2D GWAS was published as part of this thesis in 2013, incorporating published studies from Sladek 2007 to Pasquale 2013 [9]. Since this time, additional large T2D GWAS have been conducted in a range of populations. In 2013, Ma et al identified a novel T2D locus via meta-analysis of glucose traits in Han Chinese samples, with a SNP (rs10229583) near *PAX4* being associated with elevated fasting plasma glucose, impaired beta cell function in controls, and an earlier age at diagnosis for cases [40]. The first T2D GWAS conducted in an Arab population discovered a novel T2D variant at *GABRA4*, which is involved in insulin secretion [41]. A Japanese GWAS identified three novel variants in *MIR129-LEP*, *GPSM1* and *SLC16A13* [42] while a GWAS of young-

onset T2D in American Pima Indians identified a risk variant in *DNER*, which regulates the expression of Notch signalling pathway genes [43]. A common novel T2D variant in *SLC16A11* was found in Mexican and Latin American populations; it was common in Native Americans (minor allele frequency ~50%), but less common in East Asians (MAF~10%) and rare in European and African populations [44].

A notable study published in 2014 involved trans-ethnic GWAS meta-analyses of T2D performed in combined European, East Asian, South Asian, Mexican and Mexican American populations. This trans-ancestry meta-analysis identified seven new T2D susceptibility loci and improved fine-mapping of regions previously associated with the disease by utilising varying linkage disequilibrium patterns across ancestral groups [45]. Familial young-onset T2D was found to be associated with rs1408888 of *DACH1* in a Chinese population [46], with this gene having a role in pancreatic islet development and insulin secretion. Another meta-analysis performed in African Americans identified two novel loci at genome wide significance: *HLA-B* and *INS-IGF2* [47]. The year 2014 also saw the first GWAS in a Lebanese population, confirming the role of *CDKAL1* and *TCF7L2* in T2D susceptibility in Lebanese [48]. The first T2D GWAS in an Australian Aboriginal population was conducted in 2015 using 402 individuals from extended pedigrees representing a Western Australian Aboriginal group. Although none of the SNPs reached genome wide significance, common risk variants identified in other populations; *TCF7L2*, *KCNJ11*, *GABA*, *MC4R* and *IGF2BP2* were also identified in this group [49].

Another eight novel loci were identified via two large GWAS in Japanese and multi-ethnic populations published in 2016 [50, 51]. Seven SNPS reached genome wide significance in Japanese, including rs1116357 near *CCDC85A*, rs147538848 in *FAM60A*, rs1575972 near *DMRTA1*, rs9309245 near *ASB3*, rs67156297 near *ATP8B2*, rs7107784 near *MIR4686* and rs67839313 near *INAFM2*. Variants in *TOMM40-APOE* were associated with T2D in a multi-ethnic sample.

1.1.1.1 Population Genetic Differences

Considering the evidence accumulated to date, a number of T2D risk variants clearly show association across multiple populations, while some appear specific to certain population groups. Ethnicity-specific findings may partly reflect differences in the patterns of linkage disequilibrium (LD) between associated marker loci and disease, reflecting different demographic histories and population differences in evolutionary recombination [52]. Older populations such as those in Africa tend to exhibit less extensive LD due to having historically more generations in which genetic recombination events could have occurred. For this reason, such populations can be helpful for finely localizing a risk variant following an initial association finding [53].

As noted above, population frequency differences in T2D risk alleles can influence statistical power for identification. Higher risk allele frequency (RAF) corresponds with greater statistical power for a given effect size, increasing the probability of detecting a genotype-phenotype association; conversely,

lower RAF reduces power. Risk allele frequency differences also give rise to population differences in the attributable risk corresponding to an allele [29, 54].

An earlier study showed that among a range of 1,495 complex diseases, including multiple sclerosis, breast cancer and rheumatoid arthritis, T2D had the highest between-population variation in RAF for known genetic risk alleles [55]. T2D risk allele frequencies demonstrated a clear gradient matching paths of early human migration, which is consistent with the 'thrifty genotype" hypothesis [56]. This hypothesis proposes that susceptibility to obesity and T2D in some populations reflects historical, positive selection for genotypes promoting efficiency of metabolism, energy and fat storage, thus providing a survival advantage during times of nutrient shortage [57] within historical feast and famine cycles [58]. If this hypothesis does have validity, the elevated frequencies of such "thrifty" genotypes may contribute to the marked elevation in T2D prevalence now seen among certain populations [59-61]. They could also contribute to observed patterns of higher susceptibility to abdominal obesity at lower BMI and reduced muscle mass with increased insulin resistance as observed in Asian populations [62].

Among certain Asian populations, higher prevalence of T2D has also been found in those who have adopted a Western lifestyle [63-65]. In conjunction, several studies have found evidence for positive interactions between T2D genetic risk scores and "Westernized" dietary patterns characterised by increased red and processed meat intake, increased fried food consumption and reduced dietary fibre [66, 67]. Conversely, diets characterised by a low glycaemic index and high fibre intake have been shown to negatively interact with the *TCF7L2* risk variant, reducing its effect on T2D risk [68, 69]. These studies support possible contributions of gene-environment interactions to T2D risk, together with a potential model whereby interactions between recent lifestyle transitions and genetic risk factors may be contributing to the rapidly increasing prevalence of T2D in Asian populations.

1.1.1.2 Impact of Genetic Findings for T2D

Although GWAS have identified and confirmed a range of new genetic risk loci for T2D, the clinical predictive value of these loci appears low, with most variants having small effects on overall disease risk. Further, for most genetic risk loci identified, their functional role in T2D pathogenesis remains unknown. Of note however, the proteins encoded by *KCNJ11 (E23K)* and *PPARG (P12A)* – both of which have been identified by GWAS - are the therapeutic targets of two diabetes drugs; thiazolidinediones and sulphonylureas [70]. This emphasises the potential clinical value of GWAS findings.

Importantly, the combination of known genetic risk loci for T2D explains only a minority of the total genetic component of disease [71], a problem known as "missing heritability". This "missing" heritability may in part reflect the presence of additional, as yet unidentified, small-effect variants which potentially differ in frequency or effect between population groups. It is also unknown to what extent genetic risk variants are modified by environmental risk factors to further contribute to population differences in risk.

Identifying T2D risk factors in multi-ethnic populations, and better understanding their potential interactions with lifestyle risk factors, may provide insights into population differences in the incidence and prevalence of T2D.

1.2 Type 2 Diabetes in Malaysia

Malaysia, a populous Asian country, has one of the highest comparative prevalences of T2D among Asian nations [72]. In spite of this, the Malaysian population remains relatively understudied among Asian populations previously included in GWAS of T2D.

In Malaysia, the diabetes prevalence has increased almost threefold in the past three decades (Figure 1.1). The first National Health Morbidity Survey (NHMS) was conducted in 1986 and found that 6.3% of adults aged 35 years or older were living with diabetes, with the prevalence increasing to 16.6% in 2014. The total population of Malaysia was 28.3 million in 2011 and largely comprises three major ancestral groups: Malays (67%), Chinese (25%), and Indians (7%) [73].The point prevalence of T2D differs markedly between these three groups; among the three, Indians have the highest prevalence, ranging from about 25% to 28%, followed by Malays (17% to 19%), and Chinese (9% to 14%) [74]. Although environmental factors may contribute to these differences, given the shared environment among these three groups, genetic differences may also be important.

The relative importance of environmental and genetic factors in T2D have been studied in previous studies, but none of these have focused on the Malaysian population [75-78]. Thus, the current thesis aims to assess and characterize the relative contributions of genetic variation, environmental risk factors and gene by environment interaction to T2D risk in the Malaysian population.

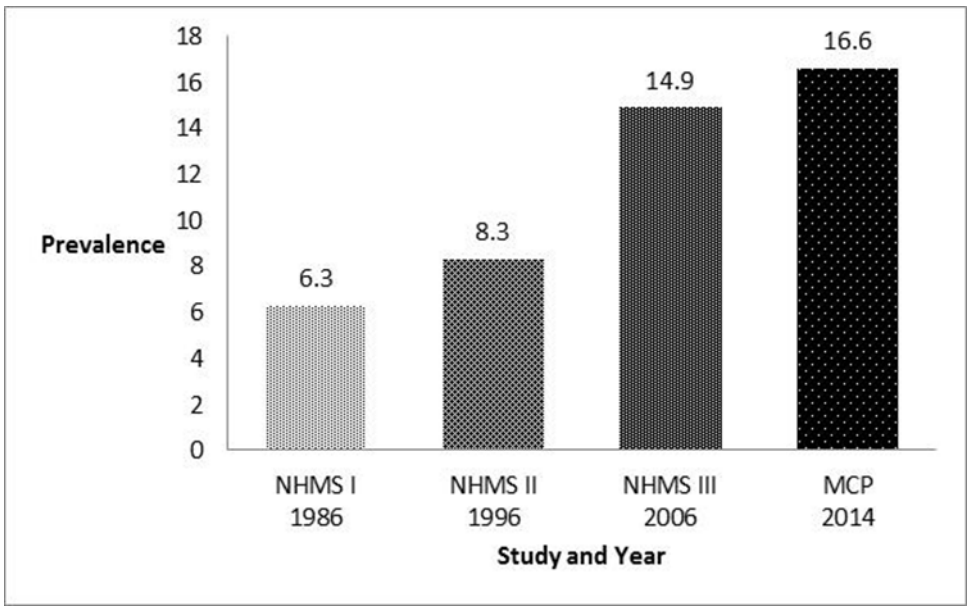


Figure 1.1: T2D prevalence trend in Malaysia

1.3 The Malaysian Cohort (MCP)

This research utilized samples of Malay, Chinese, and Indians ancestry derived from The Malaysian Cohort Project (MCP), a prospective cohort study of Malaysian participants. This cohort was collected using a mixed approach of voluntary participation (through advertisements and publicity campaigns) in urban areas together with cluster and targeted sampling for rural areas [74]. A total of 106,527 participants were recruited from April 2006 to the end of September 2012 from across Malaysia, including both urban and rural residents.

Participants from rural areas were chosen from agricultural regions listed by the Malaysian government's Federal Land Development Authority (FELDA). For urban areas, participants were recruited via publicity events held in cities, towns, government offices, private agencies and housing areas as well as via newspaper advertisements. Inclusion criteria were Malaysian citizenship with a valid identification card, being free of acute illness at the time of study entry, and giving informed consent to participate in the study. Those who refused to give consent and those with acute illnesses, including cancer, were excluded. Those with chronic disorders such as T2D were eligible for inclusion.

The MCP collected information using questionnaires, anthropometric measurement and biospecimen collection. Information collected by questionnaires comprised demographic details, occupational history, tobacco and alcohol consumption, dietary intake patterns, physical activity, reproductive history and medical history. Dietary intake was assessed using 24-hour recall and a 2-day food record. The physical activity questionnaire was adapted from the short version of the international physical activity questionnaire [79].

Anthropometric measurements were taken using a Harpenden stadiometer for height and a Seca weighing scale for weight, which were used to derive BMI. Waist and hip circumference were measured using a Seca measuring tape and used to derive waist-to-hip ratio (WHR). Body composition analysis was performed using an InBody 720 system (Biospace). Blood pressure was measured using a HEM-907 model blood pressure monitor (OMRON). An electrocardiogram was also performed. Each measurement was taken three times, with the average of the three recorded.

For biospecimen analysis, a total of 40 ml blood and 20 ml urine were collected using the UK Biobank protocol [80]. Biochemical measurements tested included fasting blood glucose, full blood count, lipid profile and renal profile. Participants were requested to fast for 8 hours prior to providing biospecimens. To ensure the quality and preservation of biospecimens in rural areas where electricity supply was absent or unreliable, a mobile laboratory was used, together with a transportation service transferring specimens from the recruitment site to the central processing site in Kuala Lumpur (UKM Medical Molecular Biology Institute: UMBI) within 24 hours. Biospecimens obtained from recruitment centres in East Malaysia were transported via air shipment within 24 hours.

1.4 Objectives

This study had four major aims:

1. To identify and highlight the importance of accounting for population diversity in genetic and non-genetic studies of T2D.
2. To assess and characterize association between environmental/lifestyle risk factors and T2D within and across the Chinese, Indian, and Malay ancestry groups in Malaysia.
3. To investigate association and the relative contribution of genetic variants to the risk of T2D in the three Malaysian population groups.
4. To assess the influence of gene-environment interactions on the risk of T2D in the three Malaysian population groups.

1.5 Research Methodology

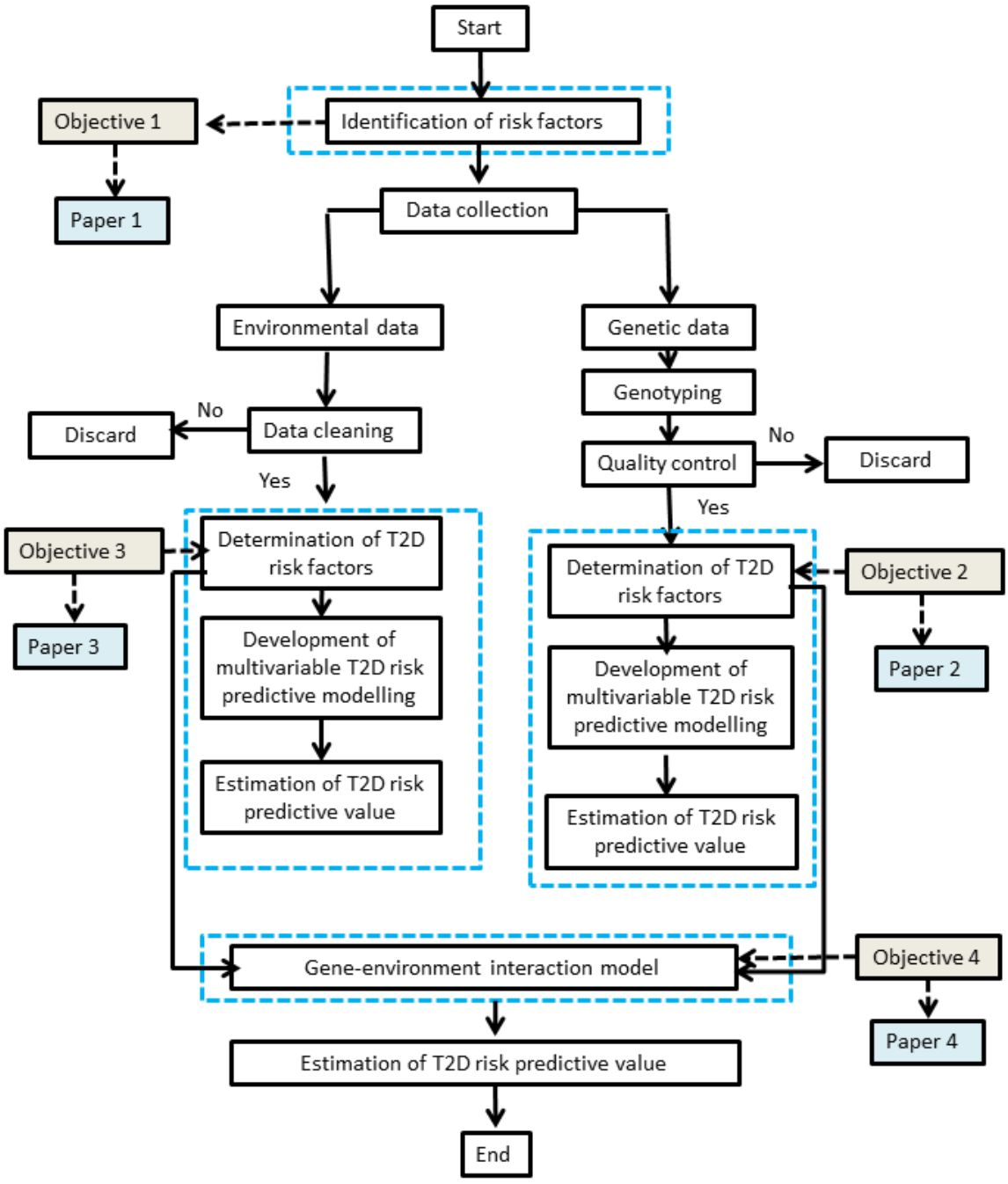


Figure 1.2: Schematic diagram showing research flow and development of scientific papers

Figure 1.2 shows the overall flow of this research and development of the four publications resulting from this research. Research methods are detailed below:

1.5.1 Search Strategy for Risk Profiles of Type 2 Diabetes

Studies investigating risk factors for type 2 diabetes, and potential population differences among these, were identified using PubMed, an online library developed and maintained by National Centre of Biotechnology Information (NCBI), at the National Institutes of Health (NIH), USA. Search terms included the keywords "racial", "ethnic", "population" or "population differences" cross referenced with the terms "type 2 diabetes", "insulin resistance", or "insulin secretion", together with "risk factors" or "risk".

For genetic risk factors, a comprehensive literature search of Genome-wide association studies (GWAS) in type 2 diabetes were performed using PubMed, together with "A Catalog of Published Genome-Wide Association Studies", an online database developed by the National Human Genome Research Institute (NHGRI) at the National Institutes of Health, USA. In the NHGRI catalogue, all studies curated under relevant categories were retrieved. These comprised categories with the following descriptions: "Type 2 Diabetes", "Type 2 Diabetes and six quantitative traits", "Type 2 diabetes and gout", "Type 2 Diabetes and other traits" and "Type 2 Diabetes and nephropathy". In parallel, the keywords "GWAS" and "Type 2 Diabetes" and "Genetics" were used to search PubMed for T2D GWAS. For studies retrieved by either approach, original articles were obtained. In addition, reference lists of all relevant articles and reviews were carefully searched for additional studies. Among these, genome-wide association studies reporting one or more single nucleotide polymorphisms (SNPs) associated with type 2 diabetes at a *P*-value < 1 x10^{-5} were retrieved, with only those associations reaching $P < 5 \times 10^{-8}$ classified as "genome-wide significant", based on this being a very widely accepted threshold. Only studies published in English were included.

1.5.2 Study Sample and Data Sources

Study samples and data were obtained from The Malaysian Cohort Project (MCP), a population-based cohort of 106,527 volunteers aged from 35 to 70 years at the time of study entry [74]. For the current research, T2D cases and controls were randomly sampled from within each of the three principal Malaysian ethnic groups: Indian, Chinese and Malay. Participants of the MCP were recruited from regions across Malaysia between April 2006 and September 2012. T2D was classified based on fasting plasma glucose (FPG) exceeding 7.5 mmol/L, with controls defined as individuals with a FPG lower than 5.5 mmol/L. A slightly higher threshold of >7.5 mmol/L was used for the current study to help ensure cases were truly diabetic and reduce the potential bias resulting from misclassification. Selecting

participants with more extreme values of a continuous trait can also help to increase power of analyses of genetic association and gene-environment interaction [81].

Controls were matched by ethnicity and randomly selected from participants with fasting blood glucose <5.5 mmol/L and no medical history of diabetes. Aside from ethnicity, controls were not matched for other characteristics. Due to selecting controls with no medical history of diabetes, controls would not expect to have known diabetes or be taking diabetic medications.

The total sample used for this research included 4077 participants, of which 1410 were Indian (708 cases, 702 controls), 1344 were Chinese (654 cases, 690 controls), and 1323 were Malay (600 cases, 723 controls). The study utilized self-reported ethnicity of the participants and their family's preceding three generations to define ethnicity. In line with the Helsinki declaration, all the relevant ethical requirements for the MCP were sanctioned by the Universiti Kebangsaan Malaysia (UKM) institutional review and ethics board. In addition, a written informed consent was also obtained by each of the participants of the study.

1.5.2.1 Genetic Data

The MetaboChip array was used to genotype samples at the UKM Medical Molecular Biology Institute in Kuala Lumpur. The custom MetaboChip array encompasses 196,725 variants from loci implicated in cardiometabolic disease traits, including T2D. This array thus offered a cost-effective, high-throughput, approach to genetic research of T2D. Illumina GenomeStudio software was used for genotype calling.

1.5.2.2 Non-genetic Data

1.5.2.2.1 Selection of environmental (non-genetic) risk factors

Environmental (non-genetic) risk factors comprising clinical, demographic and anthropometric risk factors for T2D were selected using evidence from previous studies [82-85], based on availability in the MCP study. These risk factors were communally referred as "environmental" risk factors, to distinguish them from genetic factors. They comprised age, gender, physical activity, body mass index (BMI), waist circumference (WC), waist-to-hip ratio (WHR), current smoking status, deep fried food consumption and coffee consumption. We also selected two potentially relevant, novel risk factors: sautéed food consumption and coconut milk intake, based on evidence for foods high in trans-fat being associated with cardiometabolic disease and insulin resistance [86]. Environmental risk factors were measured either using self-report questionnaires or anthropometric measurements.

1.5.2.2.2 Questionnaire-derived variables

Information related to demographic and environmental factors was collected by questionnaires and interviews at baseline [74]. Self-report questionnaires were used to measure age, gender, current smoking (yes/no), frequency of deep fried food consumption, frequency of drinking coffee and physical activity. Smoking was assessed by asking the participant whether they currently smoked or used tobacco. Dietary variables were measured by asking participants how often they had consumed foods of a particular type or prepared using specific methods in the preceding week. Physical activity was assessed using self-reported average weekly vigorous activity over the last four months using the validated International Physical Activity Questionnaire (IPAQ-M) [79].

1.5.2.2.3 Anthropometric measurement

Height, weight, body mass index (BMI), waist circumference (WC) and waist- to-hip ratio (WHR) were measured three times using a Seca or Harpenden stadiometer and averaged. BMI was measured by calculating body weight in kilograms divided by height in meters squared. WC was measured midway between the top of the hip bone and the bottom of the ribs using a tape measure. Hip circumference was measured between hips and the widest part of the buttocks.

1.5.3 Research Methods

An overview of the research methods is provided here; specific details are also included within individual research chapters.

1.5.3.1 Genetic Data

1.5.3.1.1 Genotyping and quality control

Quality control of genotype data was performed using PLINK [87, 88]. Quality control assessments and filters were performed at both the marker (SNP) and sample level. At the SNP level, markers were removed if they demonstrated minor allele frequency (MAF) <0.01, call rate (<0.95), or deviation from Hardy-Weinberg Equilibrium (HWE) in controls ($P<10^{-6}$). Samples were removed if they demonstrated missingness >0.05, outlying heterozygosity (+/- 8 SD from the mean), discrepant clinical and genotypic gender, accidental duplication or cryptic relatedness with another study sample (IBS sharing proportion >0.1875; midway between second and third-degree relatives).

1.5.3.1.2 Principal Component Analysis (PCA)

PCA is a multivariate analysis approach for reducing a set of data with a large number of variables to a small number of principal components (PCs) representing the major dimensions, or patterns of clustering between variables. Thus, the first PC represents the dimension explaining the largest amount of variability within the original data, followed by PC2 which retains the next largest amount of data variance and is orthogonal to PC1. PCA was first applied to human genetic variation by Cavalli-Sforza [89]. In large-scale genetic studies, PCA is used to identify major axes of ancestral variation in order to address and account for population stratification, a major potential source of confounding and spurious association results [90].

PCA was performed using EIGENSTRAT software and reference data from the Singaporean Genome Variation Project (SGVP) [91]. The SGVP was used due to high similarity between the Singaporean and Malaysian populations. The SGVP includes reference genotype data for 89 Singaporean Malays, 96 Singaporean Chinese and 83 Singaporean Indians. Malaysian Cohort samples not clustering with their specified ancestral group (± 6 SD from the cluster mean on the first two principal components) were removed.

After performing SNP- and sample-level quality control, including PCA, logistic regression of case-control status was performed against allelic dose across all remaining SNPs within each ancestral group. These analyses were performed with sequential adjustment for up to ten principal components (PCs) to calculate genomic inflation factors (λ_{GC}) and inform decisions about PC inclusion in candidate SNP association models, in order to minimise λ_{GC}, or test statistic inflation as a result of population stratification.

1.5.3.1.3 Selection of SNP candidates

An initial set of 188 SNPs were selected, based on showing genome-wide significant association with T2D ($P<5X10^{-8}$), as reported in a published genetic association study and the online Catalog of published genome-wide association studies [9, 92]. Of the identified 188 SNPs, 72 were available in both the Malaysian sample and MetaboChip array and also passed quality control in at least two of the three population groups in Malaysia. For loci with multiple available SNPs, a single lead SNP was selected, using the largest research study reporting association for the locus. The final set of 62 candidate SNPs were in approximate linkage equilibrium, with all pairwise squared correlation coefficients (r^2) being less than 0.5 based on linkage disequilibrium patterns in HapMap Japanese/Chinese combined reference data [93]. Based on Bonferroni correction for 62 SNPs, a multiplicity adjusted significance threshold of $\alpha = 0.05/62 = 8.06 \times 10^{-4}$ was pre-specified for association tests of individual SNPs.

1.5.3.1.4 Construction of the genetic risk score

Constructing a genetic risk score (GRS) is a statistical approach to aggregating risk effects across multiple genetic loci into a single predictive score. Such aggregation has been commonly used in genetic studies of complex traits, due to their polygenic nature and small individual effect sizes.In this study, GRS were constructed as a weighted sum of the number of risk alleles at each candidate SNP, with weights identified as the beta coefficient (log odds ratio) reported in the original publication. In the event that multiple studies had reported genome-wide significant association of a SNP, the effect estimate reported by the largest study was used. PLINK (http://pngu.mgh.harvard.edu/~purcell/plink) was used to execute a scoring algorithm to generate the GRS. Specifically, GRS ("profile scoring") was performed using the –score function, using binary format genetic input files (.bed, .bim, .fam) and a "profile.raw" file specifying the SNP ID, reference allele and score (or weight) for each allele [87, 88, 94].

1.5.3.2 Non-genetic Data

1.5.3.2.1 Data Cleaning and Coding

Data checking and cleaning were performed to ensure data quality and the plausibility of values. Categorisation of data and coding was based on approaches used in previous studies or internationally accepted criteria for each variable. Dietary frequency questions had five response choices which were categorised into 3 groups: less than once per week, 1 to 3 times per week and 4 or more times per week following a previous study [66]. Physical activity was categorised as either active or inactive using a threshold of 150 minutes per week for additional health benefit [95]. BMI was categorised as: <25 kg/m^2 (normal), 25-30 kg/m^2 (pre-obese) and >30 kg/m^2 (obese) [96-98]. WHO cut-offs for BMI were used in to allow comparison with international studies of both Asian and non-Asian populations. For WC and WHR, sex-specific cut-offs based on WHO criteria were used to derive three categories [99].

For WC these were: low risk (males: < 94cm; females: <80cm); moderate risk (males: 94-102cm; females: 80-88cm) and; high risk (males: >102 cm; females: >88cm). WHR was categorised as: low risk (< males: <0.95; females: <0.80); moderate risk (males: 0.96-1.0; females: 0.81-0.85) and; high risk (males: >1; females: >0.85).

1.5.3.2.2 Missing Data Handling by Multiple Imputation

Missing data were substantively due to the physical activity variable, resulting from a transition between two versions of physical activity questionnaires during the study. Missing data was handled by performing complete case analysis, and also via multiple imputation of missing data. Multiple imputation was performed by chained equations (MICE) with 25 cycles using STATA v11.2 (Stata Corporation, College Station, Texas) [100, 101]. MICE operates under the assumption that the missing data are Missing At Random (MAR), which means that after controlling for all of the available data included in the imputation model, any remaining missingness is assumed to be completely random (MCAR) [102]. In each cycle, missing values in each variable were imputed based on a predictive distribution derived from regression on all other variables in the imputation model.

1.5.4 Statistical Analyses

1.5.4.1 Multivariable Logistic Regression Modelling

Univariate and multivariable logistic regression was used to investigate associations between risk factors and T2D within each of the three ancestral groups separately, and also in the combined population. For each analysis, variable selection approaches based on a Change in Estimate (CIE) approach were used as previously described [103]. Multivariable models including all selected risk factors were initially fitted then the least significant risk factor (P>0.20) was removed one at a time provided the likelihood ratio P-value exceeded 0.20 and the remaining estimated coefficients (on the logit scale) of the remaining variables did not differ by more than about 10%. Any risk factor that had been removed for a particular ancestral group but retained for any other group was re-included in the final model for each group to ensure final models were comparable across ancestral groups. The model for the combination of all three ancestral groups included ethnicity as a fixed effect. Multiple logistic regression models were fitted including genetic risk factors only, non-genetic risk factors only and a combination of genetic and non-genetics risk factors, corresponding to different research questions and as described in individual chapters. The risk explained by the risk factors in each model was estimated using McFadden's pseudo R^2 and the Area Under the Receiver-Operating Characteristic Curve (AUROC), with its 95% confidence interval.

The increment in variance explained as a result of incorporating GRS to the model encompassing the environmental risk factors only was estimated using McFadden's pseudo R^2. The nested models were compared by conducting a Likelihood Ratio Chi-square test, and using De Long's test [104, 105] to compare the Area Under the Receiver-Operating Characteristic (AUC) curve. The AUC measures the predictive power and goodness of fit of logistic models, by quantitating a model's ability to distinguish between two outcome categories (here, T2D and normal). An ideal test has an AUC of 1, whereas a process of random guessing would produce an AUC of 0.5. Values of about 0.8 or greater are

often considered clinically useful. Comparisons of these statistics were performed across all ancestral groups, adjusting for ancestry as a fixed effect, and also in each group individually by using STATA 11.2 (Stata Corporation, College Station, Texas).

1.5.4.2 Multiplicative Gene-environment interaction

Based on work by Vanderweele [106], interaction was defined as the effect of one exposure on an outcome depending on the presence or absence of another exposure. Multiplicative interaction was assessed between each environmental risk factor and individual SNP, as well as the GRS, adjusting for gender and age. Interaction was assessed across all groups combined, as well as within each group individually. To supplement formal interaction tests, evidence for a gradient of GRS effects across ordered categorical strata of environmental risk factors was also assessed. The global significance of the interaction was assessed using a likelihood ratio test. In assessing the interaction between individual SNPs and each of environmental risk factors, Bonferroni correction was applied for the five environmental risk factors (age, gender, BMI, WHR, WC and physical activity, assuming high correlation between WHR and WC) and 62 SNPs, meaning the adjusted significance threshold was $0.05/ (62x5) = 1.6 \times 10^{-4}$.

1.5.4.3 Additive Gene-environment Interaction

Effect modification on the additive scale was performed by calculating the "relative excess risk due to interaction" (RERI). RERI can be calculated by using the coefficients of a logistic regression model:

$\text{logit}\,[P(D=1|G=g,E=e)] = \beta 0 + \beta 1g + \beta 2e + \beta 3eg$

where g and e represent individual genetic and environmental risk factors, with the RERI defined as:

$RERI \approx OR11 - OR10 - OR01 + 1 = \exp(\beta 1 + \beta 2 + \beta 3) - \exp(\beta 1) - \exp(\beta 2) + 1$

Values of RERI > 0 indicate a positive additive interaction; while RERI < 0 indicates negative additive interaction. Results were presented based on recommendations by Knol and VanderWeele [107]. All analyses were performed using STATA v11.2 (Stata Corporation, College Station, Texas).

CHAPTER 2: THE ARCHITECTURE OF RISK FOR TYPE 2 DIABETES: UNDERSTANDING ASIA IN THE CONTEXT OF GLOBAL FINDINGS

Abdullah N, Attia J, Oldmeadow C, Scott RJ, Holliday EG.

International Journal of Endocrinology.2014;2014:593982.doi:10.1155/2014/593982. Epub 2014 Mar 13

2.1 Statement of Co-authors

"As co-authors of the paper:
Abdullah N, Attia J, Oldmeadow C, Scott RJ, Holliday EG. The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings. International Journal of Endocrinology. 2014;2014:593982. doi: 10.1155/2014/593982. Epub 2014 Mar 13., we confirm that Noraidatulakma Abdullah contributed to this publication by performing the analysis, interpreting the result and writing the manuscript."

Co-Authors:

Dr Elizabeth Holliday
Signed
Date: 31st July 2017

Professor John Attia
Signed
Date: 1st August 2017

Dr Christopher Oldmeadow
Signed
Date: 1st August 2017

Professor Rodney Scott
Signed
Date: 2/8/2017

Faculty Assistant Dean (Research Training)
Professor Robert Callister

Signed
Date: 3/08/2017

2.2 Summary of Publications 1

Introduction

The prevalence of Type 2 Diabetes (T2D) is rising worldwide and particularly rapidly in Asian countries. Factors driving the epidemic of T2D in Asian countries remain unclear, as does the factors responsible for disease prevalence differences between different populations sharing the same geographical area and environment. Such knowledge may help to effectively manage T2D in multi-ethnic Asian populations. Malaysia has the highest comparative prevalence of T2D in Asian countries, and has a diverse population representing several distinct ancestral groups (Malay, Indian, Chinese). However T2D in Malaysia has been relatively understudied.

Results from previous studies suggest population differences in the prevalence of both genetic and non-genetic risk factors among global population groups. Recent genome-wide association studies (GWAS) in diverse populations have provided insights into the genetic architecture of T2D, as well as the potential contribution of genetic factors to population differences in risk.

This chapter comprised a comprehensive literature review to catalogue previously identified genetic and non-genetic risk factors for T2D, both in the global context, and among Asian population groups. The aim was to identify genetic and non-genetic risk factors for subsequent study in our Malaysian sample, to characterise and compare the contribution of individual risk factors to T2D prevalence among major Malaysian population groups.

Hindawi Publishing Corporation
International Journal of Endocrinology
Volume 2014, Article ID 593982, 21 pages
http://dx.doi.org/10.1155/2014/593982

Review Article

The Architecture of Risk for Type 2 Diabetes: Understanding Asia in the Context of Global Findings

Noraidatulakma Abdullah,[1,2] John Attia,[3,4] Christopher Oldmeadow,[3,4] Rodney J. Scott,[1,5] and Elizabeth G. Holliday[3,4]

[1] *School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Newcastle, NSW 2308, Australia*
[2] *UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia*
[3] *Clinical Research Design, IT and Statistical Support (CReDITSS) Unit, Hunter Medical Research Institute, Newcastle, NSW 2305, Australia*
[4] *Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, Faculty of Health, University of Newcastle, Newcastle, NSW 2305, Australia*
[5] *Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW 2305, Australia*

Correspondence should be addressed to Elizabeth G. Holliday; liz.holliday@newcastle.edu.au

Received 24 September 2013; Accepted 30 January 2014; Published 13 March 2014

Academic Editor: Mario Maggi

The prevalence of Type 2 diabetes is rising rapidly in both developed and developing countries. Asia is developing as the epicentre of the escalating pandemic, reflecting rapid transitions in demography, migration, diet, and lifestyle patterns. The effective management of Type 2 diabetes in Asia may be complicated by differences in prevalence, risk factor profiles, genetic risk allele frequencies, and gene-environment interactions between different Asian countries, and between Asian and other continental populations. To reduce the worldwide burden of T2D, it will be important to understand the architecture of T2D susceptibility both within and between populations. This review will provide an overview of known genetic and nongenetic risk factors for T2D, placing the results from Asian studies in the context of broader global research. Given recent evidence from large-scale genetic studies of T2D, we place special emphasis on emerging knowledge about the genetic architecture of T2D and the potential contribution of genetic effects to population differences in risk.

1. Introduction

Type 2 diabetes (T2D) is one of the top five noncommunicable diseases globally, comprising a major, growing cause of morbidity and premature death. In 2012, the International Diabetes Federation (IDF) estimated that 371 million people worldwide were living with diabetes, of which about half live in South Asia, the Western Pacific, and Eastern Mediterranean regions [1]. Asia is now the epicenter of an escalating diabetes epidemic, chiefly due to population growth and ageing in India and China. Projections suggest that by 2030, more than 60% of worldwide diabetes cases will come from Asia [2, 3], with the vast majority of these being Type 2 diabetes (T2D) [4]. T2D has an enormous economic, psychosocial, and physical impact on individuals, their families, and communities, both directly and indirectly. The direct economic burden of T2D includes both recorded expenditure by health services and unrecorded costs borne by patients and their families. Indirect costs such as loss of productivity and disability are also substantial and may match or surpass direct costs. The proportion of worldwide disability-adjusted life years (DALYs) due to T2D has soared in recent decades, rising from 43% in 1990 to 54% in 2010 [5]. Temporary and permanent disability, excess morbidity, and premature death are the consequences of T2D vascular complications, including cardiovascular disease, retinopathy (blindness), nephropathy (kidney failure), and neuropathy (nerve problems) which can lead to amputation. Intangible costs due to psychosocial effects on quality of life, diminished contribution to family tasks, and reduced income of

care-giving family members are also likely substantial but difficult to assess. The enormous, growing global burden of T2D—particularly in Asia—is now viewed as a crisis by the World Health Organisation (WHO) and the United Nations (UN) [6]. There is a major worldwide push to decrease the prevalence and impact of T2D by identifying risk factors, both genetic and nongenetic. Explaining the distribution and variation of T2D susceptibility across Asia will be vital for reducing the global burden of disease, due to the demographic, cultural, and genetic heterogeneity of Asian populations, and T2D risk factor profiles between these populations [7–10].

2. Epidemiology

2.1. Burden of the Disease. The vast majority of T2D (about 80%) occurs in low- and middle-income countries (LMICs), with India and China providing the largest absolute contributions. The prevalence of T2D is also rising most swiftly in LMICs [6], particularly in Asian countries experiencing rapid economic growth (Figure 2). However, there are disparities in T2D prevalence among Asian populations; Asians from the Indian subcontinent (India, Pakistan and Bangladesh) have the highest prevalence (15.9% to 24.9%), with intermediate prevalence in Malays (11.4%to 16.9%) and reduced prevalence in Chinese (6.4% to 13.8%) [11–13]. These risk profile differences may reflect population differences in T2D risk due to ethnicity-specific diet and lifestyle, body composition, genetic effects, or gene-environment interactions, as discussed further in the sections below.

2.2. Pathophysiology. The pathogenesis of Type 2 diabetes (T2D) involves deficient insulin secretion by pancreatic β-cells, and diminished insulin effectiveness in target tissues (insulin resistance) T2D aetiology differs from that of Type 1 diabetes (T1D), in which there is absolute insulin deficiency due to the destruction of insulin-producing β-cells [14]. T2D represents 90% of all diabetes cases worldwide [4]. Impaired insulin secretion and insulin action led to an accumulation of glucose in the blood (hyperglycaemia), with adverse effects on health. Clinical features of hyperglycaemia and T2D include excessive excretion of urine (polyuria), thirst (polydipsia), constant hunger, weight loss, vision change, and fatigue [15]. These symptoms may occur suddenly but are often less marked, and T2D patients may be unaware of their illness for several years until further complications develop.

2.2.1. Insulin Resistance. Glucose homeostasis depends upon a highly regulated feedback system comprising both insulin-secreting β-cells and insulin-sensitive target tissues. The function of either component—while accounting for the associated homeostatic response of the other—can be evaluated using Homeostasis Model Assessment (HOMA) [16]. Studies assessing insulin resistance using HOMA (HOMA-IR) report continental differences in the relative contribution of insulin deficiency versus insulin resistance to T2D. Compared to healthy European-ancestry participants matched for age and body mass index (BMI), Asian Indian individuals exhibit higher insulin resistance [17] and a greater contribution of insulin resistance—relative to insulin secretion—to T2D pathogenesis [18]. One study evaluating insulin response to a fixed glucose load also showed that Japanese-Americans displayed an insulin response more similar to native Japanese than European-Americans, in spite of sharing a highly Westernized lifestyle with their European-American counterparts [8].

There is also variation in the predisposition to insulin resistance *between* Asian populations [19]. For several decades, it has been recognised that the highest propensity is present in Asian Indians, in whom insulin resistance contributes substantially to T2D pathogenesis [20], potentially reflecting ancestry-related predisposition to abdominal obesity [21, 22]. A recent population based study of 4,136 Chinese, Malays, and Asian Indians living in Singapore supported these findings, reporting substantially higher insulin resistance in Asian Indians, intermediate levels in Malays, and the lowest levels in Chinese ($P < 0.001$) [19]. Differences between Malays and Chinese were removed after adjusting for body mass index (BMI); the remaining additional resistance in Indians appeared to be mediated by a tendency to higher BMI and BMI-adjusted waist circumference, together with other unexplained factors [19].

Dickinson and colleagues studied postprandial hyperglycemia and insulin sensitivity after a 75 gram carbohydrate challenge in 60 lean, healthy individuals from five ethnic groups with similar age, BMI, waist circumference, and birth weight distributions. Prior to carbohydrate consumption, fasting insulin was significantly higher in South Indians and South East Asians, compared to European Caucasian, Arabic, and Chinese individuals ($P < 0.001$) [23]. Following the challenge, hyperglycemia was significantly higher in South East Asian and Chinese participants compared with European Caucasians, while Indians and South East Asians showed a 2-3-fold higher insulin response than Europeans [23]. A small Singapore-based study of 30 individuals also showed significantly reduced insulin sensitivity in South Indians compared with Chinese or European individuals matched for age, BMI, and physical activity [24].

2.2.2. Insulin Secretion. Impaired insulin secretion is associated with β-cell dysfunction that results in a reduced insulin-secretion response to rises in blood glucose after eating [25]. The insulin secretion response to various foods can be quantified using the insulin index; more complex relationship between insulin secretion and insulin sensitivity can be measured using the disposition index (DI), which is assessed by an intravenous glucose tolerance-test [26]. A recent family-based study found that a high-fat, low-carbohydrate dietary pattern contributed to obesity, insulin resistance, and reduced β-cell function [27]. This finding might be explained by increased free fatty acids (FFAs) reducing the expression of β-cell—specific transcription factors and impairing the β-cells' ability to respond to glucose with appropriate insulin secretion [28].

Similar to insulin resistance, insulin secretion also shows evidence of racial differences, being reduced in Asians

compared with Europeans. The insulin index of Asians is reduced almost 70% in the progression from impaired glucose tolerance (IGT) to T2D, whereas in Europeans the corresponding reduction is only 50% [29, 30]. A population based-cohort study of insulin resistance and β-cell function during pregnancy also found a significantly lower β-cell secretory response to pregnancy-induced insulin resistance in South Asian and East Asian women, compared to European participants with a similar level of insulin resistance [31].

2.2.3. Complications. T2D complications can be life-threatening and include cardiovascular disease, nephropathy (kidney disease), retinopathy (blindness), and neuropathy (nerve impairment). Observational studies in European American and African American population report that cardiovascular disease risk in individuals with T2D is more than double the rate in the general population [32] and 50% of people with T2D die from cardiovascular disease, primarily heart disease and stroke [33].

There is evidence for population differences in the rate of T2D complications, between Asian populations and broader continental groups. A cross-sectional study of 5,707 Chinese, Indians, and Malays showed that the population attributable risk of ischaemic heart disease related to T2D was the highest in Indians (40.9%), intermediate in Malays (27.9%), and the lowest in Chinese (11%)[34]. A cohort study found that the progression of kidney dysfunction in T2D was faster in Indo-Asian (Indian, Pakistani, and Bangladeshi) subjects - with an estimated 2-3-fold increase in the mean rate of rise of serum creatinine over a constant follow-up period—compared to European-ancestry subjects [35]. The prevalence of diabetic end stage renal disease (ESRD) has also been reported as significantly higher in Asian T2D subjects (52.6%) compared to Caucasians (36.2%) [36].

Another microvascular complication of T2D, diabetic retinopathy, represents about five percent of all cases of global blindness [37]. Visual impairment occurs as a result of long-term, accumulated damage to small blood vessels in the retina. A recent cross-sectional study conducted by The Diabetic Retinopathy in Various Ethnic groups in UK (DRIVE UK) found that South Asian T2D populations have significantly higher prevalence of diabetic retinopathy (42.3% versus 38%) and sight threatening diabetic retinopathy (10.3% versus 5.5%) compared to white Europeans [38].

Combined with reduced blood flow, neuropathy (nerve damage) in the feet increases the risk of foot ulcers, infections, poor wound healing, and poor distal circulation, eventually increasing the risk of limb amputation [39]. Due to the elevated risk of these life-threatening complications, mortality risk among people with diabetes is at least double that of individuals without diabetes [40].

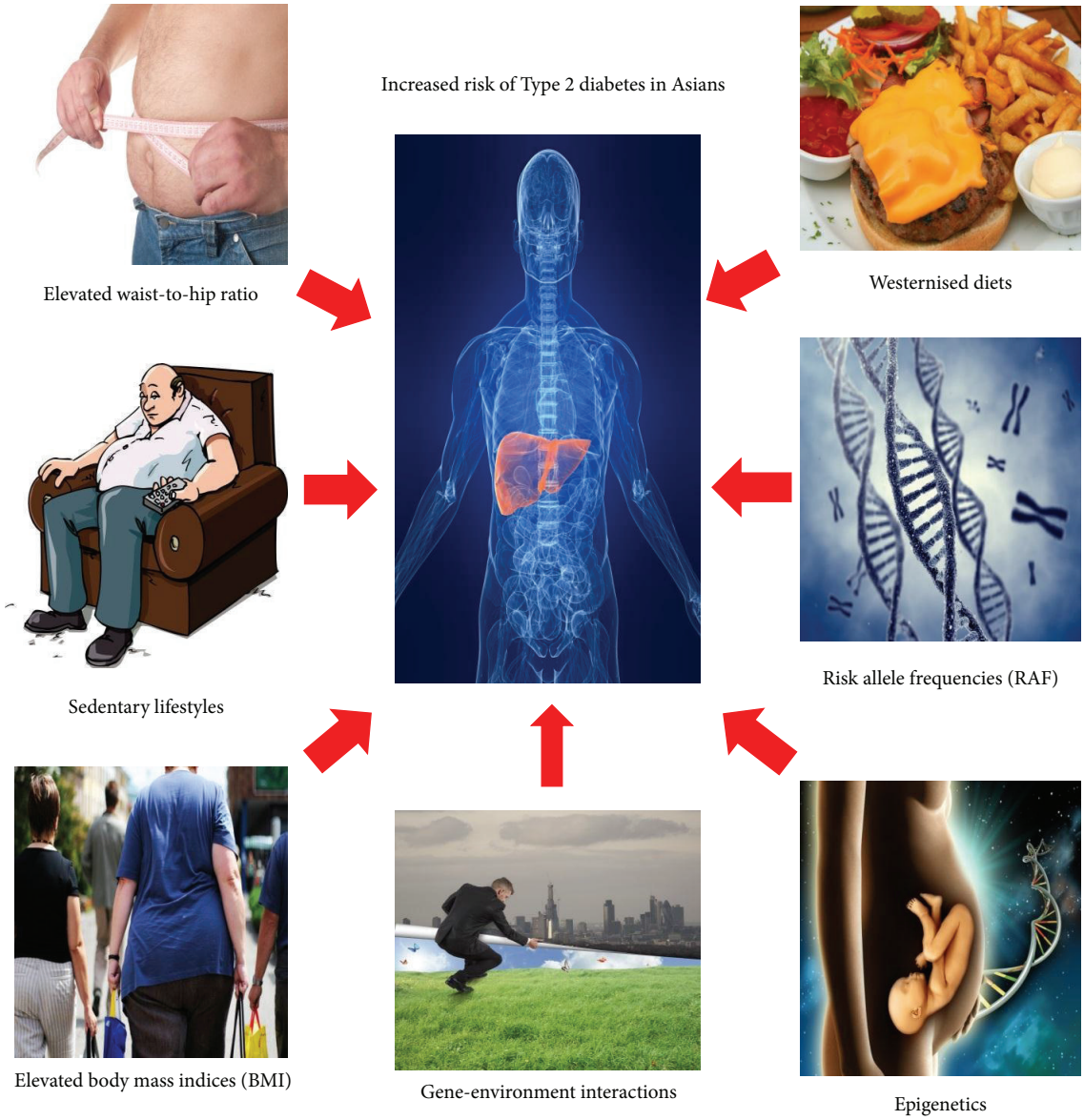
2.3. Conventional Risk Factors. A range of lifestyle and clinical factors contribute to risk of insulin resistance and T2D, including elevated body mass index (BMI), high waist-to-hip ratio (WHR), physical inactivity, and diet (Figure 1).

2.3.1. Body Mass Index (BMI) and Obesity. According to the World Health Organisation (WHO), body mass index (BMI) is a simple index of weight-for-height that can be widely used to classify overweight and obesity in adults [41]. It is defined as a person's weight in kilograms divided by the square of their height in meters (kg/m^2). Individuals with BMI greater than or equal to 30 kg/m^2 are classified as obese for international standardised comparison. Obesity elevates serum fatty acid concentrations, reducing glucose uptake and increasing fatty acid uptake by the liver, skeletal muscle, and pancreatic β-cells. Reduced glucose uptake elevates serum glucose, stimulating further insulin secretion; it is the lack of response to this secreted insulin that induces insulin resistance [42]. Continually high insulin secretion in turn produces metabolic stress in pancreatic β-cell mitochondria, inducing the release of reactive oxygen species that damage mitochondria. Over time, mitochondria lose their ability to maintain cellular processes and β-cells undergo apoptosis, irreversibly reducing insulin secretion potential [43].

Associations between BMI, percentage of body fat, and body fat distribution differ across populations, influencing the thresholds at which T2D risk increases. Asian T2D patients have lower average BMI compared to European patients [44], which might reflect higher percentage body fat in Asians (3–5% higher) than Europeans for a given BMI [45, 46]. Similarly, for a fixed body fat percentage, Asians have a 3 to 4 unit lower BMI than Europeans [45]. The body fat percentage is also different *between* Asian groups; for fixed BMI, it tends to be the highest in Indians, followed by Malays and Chinese [47]. One study also showed that among Asians, Indians have the highest prevalence of obesity (35.8% (95% CI: 32.4–39.3)), followed by Malays (32.0% (95% CI: 30.6–33.4)) and Chinese (19.7 (95% CI: 17.9–21.6)) [13]. However, due to differences in body composition, recent studies have shown that waist circumference (WC) measurement or waist-to-hip ratio (WHR) is a better predictor of T2D in Asian populations than simple BMI or body fat percentage [48, 49], since these latter measures are insensitive to differences in body fat distribution.

2.3.2. Abdominal Obesity (High Waist-to-Hip Ratio/High Waist Circumference). High waist-to-hip ratio (WHR) and waist circumference (WC), or abdominal obesity, is a major cause of insulin resistance since subcutaneous abdominal adipocytes drain their lipolytic products (free fatty acids) directly into the portal vein [50]. These free fatty acids are thought to decrease hepatic clearance of insulin and worsen systemic hyperinsulinemia [51], a precursor to T2D. Additional factors such as reduced secretion of adiponectin by adipose tissue may also contribute to the insulin-resistant state in individuals with abdominal obesity [52]. Adiponectin is an adipose tissue-specific protein that controls a number of metabolic processes, including insulin sensitivity and fatty acid oxidation [53].

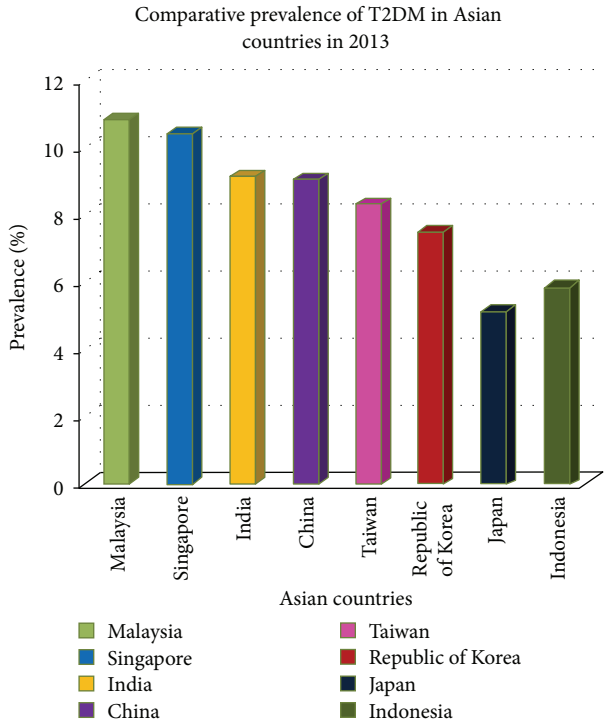
The prevalence of abdominal obesity differs between ancestral groups and seems particularly marked in certain ethnic populations such as Native Americans, African-Americans, Asians, and Pacific Islanders [54–56].



Figur e 1: Genetic and nongenetic risk factors contributing to increased Type 2 diabetes risks within Asian populations and risk differences between Asian groups.

The Multi-Ethnicity Study of Atherosclerosis found that for a given waist circumference, Chinese have the highest diabetes incidence, followed by Hispanic, African, and European-ancestry individuals [57], a finding that may be explained by higher levels of visceral adipose tissue (VAT) in Chinese compared with Europeans, at a fixed waist circumference [58]. The same study also found that South Asians have substantially higher visceral adipose tissue compared to Europeans for given waist circumference [58]. This might explain increased lipid and insulin levels observed in South Asians compared with Europeans at the same WC and/or WHR [59]. Such differences are apparent not only between Asian and other continental populations, but also among Asian populations. Among three major Asian groups, the prevalence of abdominal obesity seems to be significantly higher among Indians (61.8% (95% CI: 58.3–65.2)) compared with Malays (45.3% (95% CI: 43.8–46.8) or Chinese (40.4% (95% CI: 38.0–42.7)) [13].

2.3.3. Diet and Physical Activity. The increasing global prevalence of T2D parallels escalating obesity rates resulting from reduced physical activity, increased intake of total calories, saturated fat (especially in fast food), and sugar-sweetened beverages (SSBs) in many societies. Asian populations are undergoing a nutrition transition in conjunction with the increasing adoption of Westernized lifestyles. In India and China, for example, caloric intake from animal fat has almost doubled in recent decades [60, 61]. High consumption of red and processed meat, SSBs, and refined grains with associated low consumption of cruciferous and yellow vegetables is strongly associated with increased in T2D [62]. At the same time, physical activity has reduced in Asian populations due



Figur e 2: Comparative prevalence of Type 2 diabetes mellitus in Asian countries in 2013 (data source: http://www.idf.org/diabetesatlas/data-visualisations).

to rapid urbanization and modernization [63, 64], further increasing T2D risk.

2.3.4. Metabolic Features. Metabolic features including elevated blood pressure, hyperglycaemia, and hyperlipidaemia increase T2D risk by several-fold [65]. A recent multi-ethnic population-based survey indicated population differences in the prevalence of metabolic syndrome features, irrespective of T2D status. Indians seem to have higher levels of triglycerides and hyperglycaemia and lower HDL cholesterol, compared with Malay and Chinese [66]. These findings parallel those of a case-control study in which Indians from UK and Indians from India had higher total insulin and triglycerides and lower HDL cholesterol compared to European individuals, irrespective of shared environmental influences [22].

2.3.5. Other Factors. Other factors that have been associated with T2D risk include short sleep duration [67, 68], increasing age, which may reflect reduced exercise and muscle mass [14], history of gestational diabetes, polycystic ovary syndrome, severe mental illness, and having a family history of the disease [54]. A recent randomized, crossover study found that sleep deprivation impairs peripheral metabolic pathways, thereby reducing insulin sensitivity [69]. The loss of skeletal muscle mass with age, or sarcopenia, is also related to insulin resistance, with sarcopenia thought to cause insulin resistance and thereby increase risk of diabetes [70]. In turn, insulin resistance results in further loss of muscle strength [71]. Finally, patients with severe mental illness such as schizophrenia or bipolar disorder have 3-fold higher risk of developing T2D compared to the general population; this may result from underlying lifestyle factors, adverse effects of pharmacotherapy, and possible common genetic and/or environmentally linked pathophysiologic processes [72].

2.4. Genetic Susceptibility. In addition to conventional risk factors, family, twin and genetic studies show that T2D susceptibility has a substantial genetic component [73]. Full siblings of T2D probands have a 30–60% increased risk of disease, compared with the general population [74, 75] and children with one affected parent have a 40% lifetime risk of developing T2D, which rises to almost 70% if both parents are affected [76]. Twin studies also show higher T2D concordance in monozygotic (60–70%) compared with dizygotic twins (20–30%) [77–79].

The proportion of trait variance due to additive genetic effects is termed "heritability" and can be formally estimated from twin studies. Twin study heritability estimates are on the order of 30–70% for T2D and about 60% for impaired glucose tolerance (IGT) [80, 81]. Twin studies also demonstrate a substantial genetic component for quantitative phenotypes related to glucose homeostasis, with heritability estimates of 75–85% for *in vivo* insulin secretion, ~50% for peripheral insulin sensitivity, and ~50% for glucose metabolism [82].

Population differences in T2D pathophysiologic and risk factor profiles have been discussed in previous sections. It has been suggested that such differences may partly reflect population differences in the frequency of particular genetic risk factors and/or population-specific interactions between genetic and environmental factors [83].

2.5. Methods of Gene Identification for Common Complex Disease. Observed patterns of T2D inheritance, combined with the results of recent large-scale genetic studies, suggest that the genetic component of T2D is complex, involving multiple genetic variants of individually small effect (polygenic model) [84]. There have been three main approaches employed to identify genetic risk variants for such common complex disorders: linkage studies, candidate gene association studies (CGAS), and, more recently, genome-wide association studies (GWAS).

2.5.1. Linkage Studies. Familial linkage studies seek to identify broad genomic regions harboring disease risk variants by tracking disease and genetic marker segregation through multiple generations of families. Familial linkage studies are challenging for disorders with advanced age at onset, as parents may no longer be alive. Further challenges include difficulty in collecting accurate genealogical information and genetic (locus) heterogeneity, meaning that a particular risk locus contributes to disease in only a subset of families [85]. More broadly, this approach is limited by low power for common variants of small effect [86] and its inability to precisely localise underlying risk variants [87]. Earlier linkage studies found four (4) genetic loci linked with T2D; *CAPN10* [88],

ENPP1 [89], *HNF4A* [90], and *ACDC* (*ADIPOQ*) [91]. However, only the *HNF4A* locus has been confirmed by recent large-scale genome-wide association studies (GWAS) [92].

HNF4A, together with the related locus, *HNF1A* and also *GCK* also account for up to 80% of rare monogenic forms of diabetes. These diabetes cases present as familial, young onset, noninsulin dependent diabetes mellitus (maturity onset diabetes of the young or MODY) and are inherited in a Mendelian dominant pattern [75]. Unlike common polygenic forms of T2D, these monogenic forms require only one pathogenic genetic variant to produce disease.

2.5.2. Candidate Gene Association Studies (CGAS). In contrast to linkage studies, the candidate gene approach searches for associations between common genetic variants and disease, restricting the search region to prespecified genes of interest. Candidate genes are typically selected based on *a priori* knowledge or hypotheses reflecting the gene's presumed biological role in disease [93]. The most common study design is the case control study; for a particular genetic variant, this involves comparing the frequency of genetic alleles between individuals with and without disease, aiming to identify alleles that are associated with disease status [87]. Although a mainstay of the initial era of disease gene mapping, the candidate gene approach has been limited by small sample sizes, restriction of hypotheses to known biology, and an inability to replicate many results [94]. While candidate gene studies have reported numerous variants as beeing associated with T2D [75], just three loci, *PPARG* [95], *KCNJ11* [96], and *TCF7L2* [97], have been robustly confirmed by recent GWAS [74, 98, 99]. We note that the *TCF7L2* association study was informed by prior genome-wide linkage study showing suggestive linkage between T2D and the 10q genomic region harboring *TCF7L2* [97].

2.5.3. Genome-Wide Association Studies (GWAS). Within the last five years, genome-wide association studies (GWAS) have emerged as the method of choice for identifying common genetic variants associated with complex disease. GWAS were facilitated by completion of the Human Genome Project in 2003, the International HapMap Project in 2005 that catalogued millions of common single nucleotide polymorphisms (SNPs), and the parallel development of high throughput genotyping arrays. Single nucleotide polymorphisms (SNPs) are DNA sequence variants in which a single nucleotide differs between individuals. SNPs have a low historical mutation rate, are amenable to high throughput genotyping, and are distributed abundantly across all 22 autosomes and the sex chromosomes. Typically several million variants are screened genome-wide; appropriate adjustment of the prespecified significance level is thus necessary to avoid an increase in false positive results. Based on patterns of human genomic correlation, Bonferroni correction for one million independent tests is the accepted approach, with variants required to reach a pointwise P value $< 5 \times 10^{-8}$ (or 0.05/1,000,000) to be declared "genome-wide significant" [100]. Due to this stringent significance level, very large sample sizes are necessary to identify associations of modest effect, which is often achieved via international collaboration and the formation of consortia. The existence of such collaborations also facilitates rapid replication of findings in independent samples, a requirement for publication.

It has been eight years since the first notable GWAS finding in 2005, identifying a common allele of large effect associated with age-related macular degeneration. The year 2007 was coined the "Year of GWAS", due to the explosion of GWAS publications in that year. From 2005 to September 2013, there have been more than 1,600 GWAS published reports for a range of human diseases and traits, with an online catalogue established by the National Human Genome Research Institute at the US National Institutes of Health to collate major findings (http://www.genome.gov/gwastudies/). Although complicated and costly, GWAS have successfully identified thousands of genetic loci associated with common complex diseases under the common disease common variant (CDCV) hypothesis.

2.6. Genome-Wide Association Studies of T2D. The first T2D GWAS was published in 2007 [99], and as of September 2013, there were more than 40 publications on T2D and its complications listed in the online catalogue of published genome-wide association studies (http://www.genome.gov/gwastudies/). At the time of writing, the catalogue describes over 100 individual SNPs showing genome-wide significant association ($P < 5 \times 10^{-8}$) with T2D and related metabolic traits across diverse populations (Table 1) and over 60 SNPs showing suggestive association ($P < 1 \times 10^{-5}$) (Table 2). This section will provide a review of T2D GWAS findings to date.

The first T2D GWAS was conducted in European-ancestry participants 2007 by Sladek and colleagues [99], using a discovery sample of 600 cases and 600 controls and a separate European replication sample of nearly 3,000 cases and 3,000 controls. This small study of early onset T2D reported T2D-associated variants in three novel susceptibility genes: *TCF7L2* and *HHEX/IDE* which are associated with β-cell function [101] and *SLC30A8*, encoding a zinc transporter highly expressed in pancreatic islets [102]. Several months later, four additional European studies [74, 98, 103, 104] confirmed association of variants at these loci and identified additional associated variants in *IGF2BP2*, associated with β-cell dysfunction [105], and *CDKN2A/CDKN2B* and *CDKAL1*, which are both associated with β-cell development [105, 106]. During this time, variants in the *FTO* (fat mass and obesity associated) gene were also identified with important effects on obesity and hence, indirectly, T2D [107, 108]. Interestingly, as the effect of *FTO* variants on T2D is only via obesity, the *FTO* locus was not identified in T2D GWAS using cases and controls matched for BMI. Two of these early publications also showcased the output of international consortia: The UK-based Wellcome Trust Case Control Consortium (WTCCC) and the USA-based Diabetes Genetics Initiative (DGI), highlighting the benefits of large-scale collaboration in the GWAS era.

Since these initial GWAS had modest power to detect variants with modest effects on disease risk, follow-up studies

Table 1: Published SNPs associated with Type 2 diabetes mellitus at genome-wide significance ($P < 5 \times 10^{-8}$).

Number	SNP (allele)[1,2]	Mapped gene(s)	Region[3]	Disc[4] Pop	Rep[5] Pop	RAF[6]	RAF_{EU}[7]	RAF_{EA}[8]	RAF_{SA}[9]	RAF_{AF}[10]	OR (95% CI)	*P* value
1	rs10923931 (T) [109]	*NOTCH2*	1p12	EUR		0.11	0.09	0.03	0.22	0.39	1.13 (1.08–1.17)	4×10^{-8}
2	rs340874 (C) [155]	*LINC00538-PROX1*	1q32.3	EUR		0.52	0.56	0.32	0.55	0.08	NR	7×10^{-12}
3	rs243021 (A) [116]	*FLJ30838-MIR4432*	2p16.1	EUR	EUR	NR	0.48	0.64	0.48	0.45	1.08 (1.06–1.10)	3×10^{-15}
4	rs7578597 (T) [109]	*THADA*	2p21	EUR		0.90	0.88	1.00	0.83	0.67	1.15 (1.10–1.20)	1×10^{-9}
5	rs780094 (C) [155]	*GCKR*	2p23.3	EUR		0.62	0.61	0.43	0.80	0.88	NR	6×10^{-38}
6	rs7560163 (C) [148]	*RND3-FABP5P10*	2q23.3	AA		0.86	1.00	0.85	0.96	0.89	1.33 (1.19–1.49)	7×10^{-9}
7	rs7593730 (C) [118]	*RBMS1*	2q24.2	EUR		0.78	0.83	0.81	0.79	0.59	1.11 (1.08–1.16)	4×10^{-8}
8	rs3923113 (A) [92]	*EIF3EP3-SNORA70F*	2q24.3	SA		0.74	0.59	0.84	0.76	0.23	1.09 (1.06–1.13)	1×10^{-8}
9	rs13389219 (C) [131]	*GRB14-COBLL1*	2q24.3	EUR		0.60	0.56	0.79	—	0.15	1.07 (1.05–1.10)	1.0×10^{-8}
10	rs560887 (C) [155]	*G6PC2*	2q31.1	EUR		0.70	0.67	0.98	0.85	1.00	NR	9×10^{-218}
11	rs7578326 (A) [116]	*LOC646736*	2q36.3	EUR		NR	0.65	0.87	0.85	0.60	1.11 (1.08–1.13)	5×10^{-20}
12	rs2943641 (C) [115]	*NYAP2-MIR5702*	2q36.3	EUR	EUR	0.63	0.63	0.94	0.74	0.76	1.19 (1.13–1.25)	9×10^{-12}
13	rs4607103 (C) [109]	*ADAMTS9-AS2*	3p14.1	EUR		0.76	0.81	0.56	0.48	0.71	1.09 (1.06–1.12)	1×10^{-8}
14	rs831571 (C) [127]	*PSMD6-PRICKLE2-AS1*	3p14.1	EA		0.61	0.77	0.65	0.83	0.84	1.09 (1.06–1.12)	8×10^{-11}
15	rs11708067 (A) [155]	*ADCY5*	3q21.1	EUR		0.78	0.77	0.99	0.77	0.89	NR	7×10^{-22}
16	rs11920090 (T) [155]	*SLC2A2*	3q26.2	EUR		0.87	0.86	0.98	—	0.68	NR	8×10^{-13}
17	rs1470579 (C) [116]	*IGF2BP2*	3q27.2	EUR	SA, EA, PS	NR	0.30	0.24	0.50	0.86	1.14 (1.09–1.19)	2×10^{-9}
18	rs6769511 (C) [114]	*IGF2BP2*	3q27.2	EA		0.32	0.30	0.24	0.50	0.84	1.23 (1.15–1.33)	1×10^{-9}
19	rs4402960 (T) [98]	*IGF2BP2*	3q27.2	EUR	EA	0.30	0.30	0.22	0.49	0.54	1.14 (1.11–1.18)	9×10^{-16}
20	rs16861329 (G) [92]	*ST6GAL1*	3q27.3	SA		0.75	0.88	0.80	0.76	0.94	1.09 (1.06–1.12)	3×10^{-8}
21	rs1801214 (T) [116]	*WFS1*	4p16.1	EUR		NR	0.12	0.20	0.24	0.06	1.13 (1.08–1.18)	3×10^{-8}
22	rs4689388 (T) [115]	*WFS1*	4p16.1	EUR	EUR	0.57	0.67	0.97	0.68	0.72	1.16 (1.10–1.21)	1×10^{-8}
23	rs7656416 (C) [138]	*CTBP1-AS1-MAEA*	4p16.3	EA		0.68	—	0.70	—	0.90	1.15 (1.10–1.21)	1×10^{-8}
24	rs6815464 (C) [127]	*MAEA*	4p16.3	EA		0.58	—	0.52	—	0.92	1.13 (1.10–1.16)	2×10^{-20}
25	rs459193 (G) [131]	*ANKRD55-MAP3K1*	5q11.2	EUR		0.70	0.78	0.51	—	0.64	1.08 (1.05–1.11)	6.0×10^{-9}
26	rs4457053 (G) [116]	*ZBED3-AS1*	5q13.3	EUR		NR	0.26	0.04	—	—	1.08 (1.06–1.11)	3×10^{-12}
27	rs1535500 (T) [127]	*KCNK16; KCNK17*	6p21.2	EA		0.42	0.47	0.44	—	0.95	1.08 (1.05–1.11)	2×10^{-8}
28	rs9470794 (C) [127]	*ZFAND3*	6p21.2	EA		0.27	0.12	0.34	0.14	0.19	1.12 (1.08–1.16)	2×10^{-10}
29	rs10440833 (A) [116]	*CDKAL1*	6p22.3	EUR	EUR	NR	0.25	0.39	—	0.21	1.25 (1.20–1.31)	2×10^{-22}
30	rs4712523 (G) [144]	*CDKAL1*	6p22.3	EA	EUR	0.41	0.34	0.41	0.24	0.68	1.27 (1.21–1.33)	7×10^{-20}
31	rs10946398 (C) [74]	*CDKAL1*	6p22.3	EUR		0.32	0.34	0.40	0.22	0.67	1.16 (1.10–1.22)	1×10^{-8}
32	rs6931514 (G) [109]	*CDKAL1*	6p22.3	EUR		NR	0.28	—	0.22	0.22	1.25 (1.17–1.33)	1×10^{-11}
33	rs7754840 (C) [103]	*CDKAL1*	6p22.3	EUR	EA	0.31	0.34	0.40	0.22	0.67	1.12 (1.08–1.16)	4×10^{-11}
34	rs7756992 (G) [170]	*CDKAL1*	6p22.3	EUR		0.26	0.28	0.50	0.24	0.58	1.20 (1.13–1.27)	8×10^{-9}
35	rs7766070 (A) [171]	*CDKAL1*	6p22.3	EUR		0.27	0.25	0.40	—	0.19	1.21 (1.14–1.28)	6×10^{-11}
36	rs1048886 (G) [124]	*C6orf57*	6q13	SEA (I)		0.18	0.15	0.09	0.17	0.34	1.54 (1.32–1.80)	3×10^{-8}
37	rs4607517 (A) [155]	*GCK-YKT6*	7p13	EUR		0.16	0.20	0.22	0.12	0.06	NR	7×10^{-92}
38	rs849134 (A) [116]	*JAZF1*	7p15.1	EUR		NR	0.54	0.81	—	0.77	1.13 (1.09–1.18)	3×10^{-9}
39	rs864745 (T) [109]	*JAZF1*	7p15.1	EUR		0.50	0.49	0.77	0.79	0.77	1.10 (1.07–1.13)	5×10^{-14}
40	rs2191349 (T) [155]	*DGKB-AGMO*	7p21.2	EUR		0.52	0.48	0.71	—	0.58	NR	3×10^{-44}
41	rs10229583 (G) [172]	*FSCN3-PAX4*	7q32	EA		0.83	0.74	0.82	0.66	0.72	1.14 (1.09–1.19)	2×10^{-10}
42	rs6467136 (G) [127]	*ZNF800-GCC1*	7q32.1	EA		0.79	0.50	0.76	0.57	0.75	1.11 (1.07–1.14)	5×10^{-11}

Table 1: Continued.

Number	SNP (allele)[1,2]	Mapped gene(s)	Region[3]	Disc[4] Pop	Rep[5] Pop	RAF[6]	RAF_{EU}[7]	RAF_{EA}[8]	RAF_{SA}[9]	RAF_{AF}[10]	OR (95% CI)	*P* value
43	rs972283 (G) [116]	*KLF14-MIR29A*	7q32.3	EUR		NR	0.55	0.67	—	0.94	1.07 (1.05–1.10)	2×10^{-10}
44	rs516946 (C) [131]	*ANK1*	8p11.21	EUR		0.76	0.81	0.82	0.85	0.88	1.09 (1.06–1.12)	2.5×10^{-10}
45	rs515071 (G) [138]	*ANK1; MIR486*	8p11.21	EA		0.79	0.81	0.79	0.85	0.85	1.18 (1.12–1.23)	1×10^{-8}
46	rs896854 (T) [116]	*TP53INP1*	8q22.1	EUR		NR	0.44	0.32	0.40	0.75	1.06 (1.04–1.09)	1×10^{-9}
47	rs3802177 (G) [116]	*SLC30A8*	8q24.11	EUR		NR	0.76	0.53	0.78	0.94	1.15 (1.10–1.21)	1×10^{-8}
48	rs11558471 (A) [155]	*SLC30A8*	8q24.11	EUR		0.31	0.75	0.53	0.78	0.94	NR	3×10^{-11}
49	rs13266634 (C) [144]	*SLC30A8*	8q24.11	EUR	EA	0.57	0.76	0.53	0.78	0.94	1.22 (1.16–1.28)	2×10^{-14}
50	rs10965250 (G) [116]	*CDKN2B-AS1-DMRTA1*	9p21.3	EUR	EUR	NR	0.80	0.58	—	1.00	1.20 (1.13–1.27)	1×10^{-10}
51	rs1333051 (A) [149]	*CDKN2B-AS1-DMRTA1*	9p21.3	HIS		NR	0.84	0.85	0.93	0.91	1.22 (1.15–1.30)	6×10^{-10}
52	rs2383208 (A) [144]	*CDKN2B-AS1-DMRTA1*	9p21.3	EA	EA, SA	0.55	0.79	0.59	0.91	0.87	1.34 (1.27–1.41)	2×10^{-29}
53	rs10811661 (T) [98]	*CDKN2B-AS1-DMRTA1*	9p21.3	EUR		0.85	0.80	0.56	0.91	0.98	1.20 (1.14–1.25)	8×10^{-15}
54	rs7018475 (G) [173]*	*CDKN2B-AS1-DMRTA1*	9p21.3	EUR		NR	0.27	0.37	0.38	0.22	1.35 (1.18–1.55)	3×10^{-8}
55	rs17584499 (T) [117]	*PTPRD*	9p24.1	EA		0.06	0.23	0.11	0.25	0.03	1.57 (1.36–1.82)	9×10^{-10}
56	rs10814916 (C) [146]	*GLIS3*	9p24.2	EA		0.44	0.57	0.45	0.62	0.67	1.11(1.08–1.15)	6×10^{-12}
57	rs7041847 (A) [127]	*GLIS3*	9p24.2	EA		0.41	0.55	0.46	0.65	0.96	1.10 (1.07–1.13)	2×10^{-14}
58	rs7034200 (A) [155]	*GLIS3*	9p24.2	EUR		0.49	0.54	0.30	—	0.60	NR	1×10^{-12}
59	rs13292136 (C) [116]	*KRT18P24-CHCHD2P9*	9q21.31	EUR	EUR	NR	0.93	0.91	—	0.90	1.11(1.07–1.15)	3×10^{-8}
60	rs2796441 (G) [131]	*TLE1-FAM75D5*	9q21.32	EUR		0.57	0.61	0.40	0.55	0.90	1.07 (1.05–1.10)	5.4×10^{-9}
61	rs10906115(A) [120]	*CDC123-MIR4480*	10p13	EA		0.57	0.64	0.64	0.58	0.76	1.13 (1.08–1.18)	1×10^{-8}
62	rs11257655 (T) [146]	*CDC123-MIR4480*	10p13	EA		0.56	0.26	0.58	0.23	0.27	1.15 (1.10–1.20)	7×10^{-9}
63	rs12779790 (G) [109]	*CDC123-MIR4480*	10p13	EUR		0.18	0.22	0.13	—	0.05	1.11(1.07–1.14)	1×10^{-10}
64	rs1802295 (A) [92]	*VPS26A*	10q22.1	SA		0.26	0.35	0.11	0.29	0	1.08 (1.05–1.12)	4×10^{-8}
65	rs12571751(A) [131]	*ZMIZ1*	10q22.3	EUR		0.52	0.53	0.55	0.50	0.50	1.08 (1.05–1.10)	1×10^{-10}
66	rs1111875 (C) [98]	*HHEX-EXOC6*	10q23.33	EUR	EA	0.52	0.58	0.34	0.41	0.88	1.13(1.09–1.17)	6×10^{-10}
67	rs5015480 (C) [171]	*HHEX-EXOC6*	10q23.33	EUR	EA	0.57	0.58	0.21	0.44	0.64	1.18 (1.14–1.23)	2×10^{-9}
68	rs10885122 (G) [155]	*ADRA2A-BTBD7P2*	10q25.2	EUR		0.87	0.90	0.92	—	0.21	NR	3×10^{-16}
69	rs4506565 (T) [104]	*TCF7L2*	10q25.2	EUR		0.32	0.30	0.03	0.30	0.46	1.36 (1.20–1.54)	5×10^{-12}
70	rs7901695 (C) [74]	*TCF7L2*	10q25.2	EUR		NR	0.28	0.03	0.29	0.46	1.37 (1.31–1.43)	1×10^{-48}
71	rs7903146 (T) [116]	*TCF7L2*	10q25.2	EUR	EA, SA, AA	NR	0.28	0.03	0.28	0.27	1.40 (1.34–1.46)	2×10^{-51}
72	rs10886471 (C) [146]	*GRK5*	10q26.11	EA		0.78	0.48	0.80	0.65	0.90	1.12 (1.08–1.16)	7×10^{-9}
73	rs11605924 (A) [155]	*CRY2*	11p11.2	EUR		0.49	0.13	0.70	0.12	0.94	NR	1×10^{-14}
74	rs7944584 (A) [155]	*MADD*	11p11.2	EUR		0.75	0.71	0.96	0.78	1.00	NR	2×10^{-18}
75	rs5215 (C) [74]	*KCNJ11*	11p15.1	EUR		NR	0.40	0.38	0.4	0.01	1.14 (1.10–1.19)	5×10^{-11}
76	rs5219 (T) [98]	*KCNJ11*	11p15.1	EUR		0.46	—	—	—	—	1.14 (1.10–1.19)	7×10^{-11}
77	rs163182 (C) [145]	*KCNQ1*	11p15.4	EA		0.34	0.25	0.37	—	0.25	1.28 (NR)	2×10^{-17}
78	rs2237895 (C) [117]	*KCNQ1*	11p15.4	EA		0.33	—	—	—	—	1.29 (1.19–1.40)	1×10^{-9}
79	rs2237892 (C) [113]	*KCNQ1*	11p15.4	EA	HIS, EA	0.61	0.92	0.67	0.99	0.90	1.40 (1.34–1.47)	2×10^{-42}
80	rs2237897 (C) [114]	*KCNQ1*	11p15.4	EA		0.34	0.95	—	—	—	1.33 (1.24–1.41)	1×10^{-16}
81	rs231362 (G) [116]	*KCNQ1; KCNQ1OT1*	11p15.5	EUR		NR	0.52	0.84	—	0.86	1.08 (1.06–1.10)	3×10^{-13}
82	rs174550 (T) [155]	*FADS1*	11q12.2	EUR		0.64	0.65	0.66	—	0.98	NR	2×10^{-15}
83	rs1552224 (A) [116]	*ARAP1*	11q13.4	EUR		NR	0.87	0.91	0.76	1.00	1.14 (1.11–1.17)	1×10^{-22}
84	rs1387153 (T) [116]	*FAT3-MTNR1B*	11q14.3	EUR	EUR	NR	0.27	0.39	0.38	0.40	1.09 (1.06–1.11)	8×10^{-15}

Table 1: Continued.

Number	SNP (allele)[1,2]	Mapped gene(s)	Region[3]	Disc[4] Pop	Rep[5] Pop	RAF[6]	RAF_{EU}[7]	RAF_{EA}[8]	RAF_{SA}[9]	RAF_{AF}[10]	OR (95% CI)	*P* value
85	rs10830963 (G) [155]	*MTNR1B*	11q14.3	EUR		0.30	0.30	0.39	—	0.04	NR	6×10^{-175}
86	rs10842994 (C) [131]	*KLHDC5*	12p11.22	EUR		0.80	0.80	0.79	0.90	1.00	1.10 (1.06–1.13)	6.1×10^{-10}
87	rs1531343 (C) [116]	*RPSAP52*	12q14.3	EUR		NR	0.12	0.08	0.19	0.40	1.10 (1.07–1.14)	4×10^{-9}
88	rs7961581 (C) [109]	*TSPAN8-LGR5*	12q21.1	EUR		0.27	0.25	0.17	0.35	0.18	1.09 (1.06–1.12)	1×10^{-9}
89	rs35767 (G) [155]	*IGF1*	12q23.2	EUR		0.85	0.88	0.65	0.71	0.55	NR	3×10^{-8}
90	rs7957197 (T) [116]	*OASL*	12q24.31	EUR		NR	0.85	1.00	—	0.86	1.07 (1.05–1.10)	2×10^{-8}
91	rs7305618 (C) [149]	*RPL12P33-HNF1A-AS1*	12q24.31	HIS		NR	0.80	0.44	0.75	0.56	1.14 (1.09–1.20)	2×10^{-8}
92	rs9552911 (G) [152]	*SGCG*	13q12.12	PS	PS	0.93	1.00	0.78	0.86	1.00	1.49 (1.30–1.72)	2×10^{-8}
93	rs1359790 (G) [120]	*NDFIP2-SPRY2*	13q31.1	EA		0.71	0.73	0.69	0.84	0.92	1.15 (1.10–1.20)	6×10^{-9}
94	rs7403531 (T) [146]	*RASGRP1*	15q14	EA		0.35	0.28	0.33	0.20	0.18	1.10 (1.06–1.13)	4×10^{-9}
95	rs7172432 (A) [119]	*C2CD4A-C2CD4B*	15q22.2	EA		0.58	0.58	0.54	0.61	0.27	1.11 (1.08–1.14)	9×10^{-14}
96	rs11071657 (A) [155]	*C2CD4A-C2CD4B*	15q22.2	EUR		0.63	0.58	0.69	—	0.94	NR	4×10^{-8}
97	rs7178572 (G) [92]	*HMG20A*	15q24.3	SA	EUR	0.52	0.68	0.40	0.44	0.40	1.09 (1.06–1.12)	7×10^{-11}
98	rs7177055 (A) [131]	*HMG20A-LINGO1*	15q24.3	EUR		0.68	0.71	0.39	0.45	0.24	1.08 (1.05–1.10)	4.6×10^{-9}
99	rs16934397 (G) [116]	*ZFAND6-FAH*	15q25.1	EUR	EUR	NR	0.64	0.08	0.55	0.41	1.06 (1.04–1.08)	2×10^{-9}
100	rs2028299 (C) [92]	*AP3S2; C15orf38-AP3S2*	15q26.1	SA		0.31	0.73	0.12	0.78	0.40	1.10 (1.07–1.13)	2×10^{-11}
101	rs8042680 (A) [116]	*PRC1; LOC100507118*	15q26.1	EUR		NR	0.26	1.00	0.59	0.98	1.07 (1.05–1.09)	2×10^{-10}
102	rs11642841 (A) [116]	*FTO*	16q12.2	EUR	EUR	NR	0.47	0.06	—	0.04	1.13 (1.08–1.18)	3×10^{-8}
103	rs8050136 (A) [98]	*FTO*	16q12.2	EUR	SA	0.38	0.46	0.14	0.25	0.45	1.17 (1.12–1.22)	1×10^{-12}
104	rs9939609 (A) [171]	*FTO*	16q12.2	EUR		0.40	0.46	0.15	0.26	0.50	1.25 (1.19–1.30)	1×10^{-20}
105	rs7202877 (T) [131]	*CTRB2-CTRB1*	16q23.1	EUR		0.89	0.89	0.80	0.95	0.85	1.12 (1.07–1.16)	3.5×10^{-8}
106	rs391300 (G) [117]	*SRR*	17p13.3	EA		0.62	0.63	0.75	0.48	0.42	1.28 (1.18–1.39)	3×10^{-9}
107	rs4430796 (G) [146]	*HNF1B*	17q12	EUR	EA	0.28	0.51	0.25	0.31	0.66	1.19 (1.13–1.25)	2×10^{-11}
108	rs8090011(G) [171]	*LAMA1*	18p11.31	EUR		0.38	0.32	0.72	—	0.79	1.13 (1.09–1.18)	8×10^{-9}
109	rs12970134 (A) [131]	*MC4R*	18q21.32	EUR		0.27	0.28	0.18	0.29	0.17	1.08 (1.05–1.11)	1.2×10^{-8}
110	rs3786897 (A) [127]	*PEPD*	19q13.11	EA		0.56	0.61	0.58	0.81	0.40	1.10 (1.07–1.14)	1×10^{-8}
111	rs6017317 (G) [127]	*FITM2-R3HDML*	20q13.12	EA		0.48	0.18	0.41	—	0.59	1.09 (1.07–1.12)	1×10^{-11}
112	rs4812829 (A) [92]	*HNF4A*	20q13.12	SA		0.29	0.16	0.45	0.29	0.08	1.09 (1.06–1.12)	3×10^{-10}
113	rs12010175 (G) [146]	*FAM58A*	Xq28	EA		0.79	0.94	0.84	0.81	0.79	1.21 (1.14–1.28)	2×10^{-9}
114	rs5945326 (A) [116]	*KRT18P48-DUSP9*	Xq28	EUR	EA	NR	0.78	0.66	—	0.84	1.27 (1.18–1.37)	3×10^{-10}

*Risk Allele and RAF not reported, but chosen based on minor allele frequency (MAF) in the population mentioned in the original publication.

Table 2: Published SNPs associated with Type 2 diabetes mellitus at suggestive significance ($P < 1 \times 10^{-5}$).

Number	SNP (allele)[1,2]	Mapped Gene(s)	Region[3]	Disc[4] Pop	Rep[5] Pop	RAF[6]	RAF_{EU}[7]	RAF_{EA}[8]	RAF_{SA}[9]	RAF_{AF}[10]	OR (95% CI)	*P* value
1	rs7542900 (C) [148]	*F3-PGBD4P7*	1p21.3	AA		0.56	0.80	0.72	0.79	0.56	1.16 (1.09–1.25)	6×10^{-6}
2	rs11765354 (A) [151]	*TGFBR3*	1p22.1	SA	All SA	0.78	0.62	0.52	0.85	0.28	1.17 (1.10–1.25)	4×10^{-6}
3	rs17045328 (G) [124]	*CR2*	1q32.2	SEA (M)		0.30	0.03	0.35	0.07	0.02	1.38 (1.20–1.59)	7×10^{-6}
4	rs6426514 (A) [152]	*RHOU*	1q42.13	PS		0.06	0.09	0.03	0.02	0.12	1.51 (1.27–1.78)	2×10^{-6}
5	rs12027542 (A) [124]	*PCNXL2*	1q42.2	SEA (M)		0.61	0.93	0.69	0.95	0.95	1.41 (1.23–1.61)	4×10^{-7}
6	rs11677370 (T)[124]	*DCDC2C*	2p25.3	SEA (I)		0.4	0.68	0.71	—	0.78	1.35 (1.19–1.53)	3×10^{-6}
7	rs6712932 (C) [174]*	*MRPS9-GPR45*	2q12.1	EUR		NR	0.34	0.22	0.32	0.28	1.52 (1.27–1.82)	6×10^{-6}
8	rs6723108 (T) [151]	*TMEM163-MIR5590*	2q21.3	SA	All SA	0.86	0.50	1	0.93	1	1.27 (1.17–1.39)	7×10^{-8}
9	rs358806 (C) [104]	*LRTM1-WNT5A*	3p14.3	EUR		0.80	0.77	0.84	0.90	0.92	1.16 (1.03–1.33)	3×10^{-6}
10	rs13081389 (A) [116]	*SYN2-GSTM5P1*	3p25.2	EUR		NR	0.95	0.98	—	1.00	1.24 (1.15–1.35)	2×10^{-7}
11	rs17036101 (G) [109]	*SYN2-GSTM5P1*	3p25.2	EUR		0.93	0.95	0.98	—	0.98	1.15(1.10–1.21)	2×10^{-7}
12	rs1801282 (C) [103]	*PPARG*	3p25.2	EUR		0.86	0.90	0.94	0.91	1.00	1.14 (1.08–1.20)	2×10^{-6}
13	rs2063640 (A) [124]	*ZPLD1-NDUFA4P2*	3q12.3	SEA (M, C, I)		0.17	0.08	0.27	0.11	0.04	1.23 (1.13–1.34)	3×10^{-6}
14	rs3773506 (C) [124]	*PLS1*	3q23	SEA (I)		0.06	0.04	0.11	0.04	0.14	1.81 (1.39–2.35)	9×10^{-6}
15	rs7630877 (A) [124]	*PEX5L*	3q26.33	SEA (C)		0.17	0.35	0.18	0.36	0.31	1.32 (1.17–1.49)	7×10^{-6}
16	rs13749 10 (T) [149]	*IGF2BP2*	3q27.2	HIS		NR	0.02	0.08	—	0.15	1.24 (1.15–1.34)	1×10^{-7}
17	rs7659604 (T) [104]	*ANXA5-TMEM155*	4q27	EUR		0.38	0.44	0.36	0.45	0.71	1.35 (1.19–1.54)	9×10^{-6}
18	rs3792615 (T) [124]	*36951*	4q32.3	SEA (I)		0.95	0.97	0.84	0.96	0.85	1.93 (1.45–2.59)	9×10^{-6}
19	rs10461617(A) [151]	*RPL26P19-MAP3K1*	5q11.2	SA	All SA	0.21	0.18	0.39	0.26	0.44	1.17 (1.09–1.25)	4×10^{-6}
20	rs12518099 (C) [115]	*MIR3660-CETN3*	5q14.3	EUR		0.23	0.23	0.39	0.31	0.23	1.16 (1.10–1.22)	7×10^{-7}
21	rs17053082 (T) [152]	*PPIGP1-SGCD*	5q33.2	PS		0.1	0.06	0.06	0.06	0.08	1.49 (1.28–1.73)	4×10^{-7}
22	rs9472138 (T) [109]	*VEGFA-C6orf223*	6p21.1	EUR		0.28	0.24	0.11	0.23	0.14	1.06 (1.04–1.09)	4×10^{-6}
23	rs3916765 (A) [171]	*MTCO3P1-HLA-DQA2*	6p21.32	EUR		0.12	0.17	0.08	0.08	0	1.21(1.12–1.31)	1×10^{-6}
24	rs9295474 (G) [124]	*CDKAL1*	6p22.3	SEA (M, C, I)		0.36	0.30	0.41	—	0.36	1.16 (1.09–1.24)	9×10^{-6}
25	rs9465871 (C) [104]	*CDKAL1*	6p22.3	EUR		0.18	0.16	0.52	0.21	0.58	1.18 (1.04–1.34)	3×10^{-7}
26	rs7769051 (A) [121]	*SNORA33-HMGB1P13*	6q23.2	AA		0.29	0.10	0.04	0.16	0.38	1.28 (1.16–1.42)	2×10^{-6}
27	rs642858 (A) [124]	*ATP5F1P6-MIR3668*	6q24.1	SEA (I)		0.40	0.25	0.40	0.36	0.13	1.35 (1.19–1.53)	2×10^{-6}
28	rs6930576 (A) [121]	*SASH1*	6q24.3	AA		0.28	0.34	0.18	0.44	0.24	1.31 (1.18–1.45)	7×10^{-7}
29	rs741301 (C) [175]*	*ELMO1*	7p14.2	EA		NR	0.32	0.32	0.43	0.67	2.67 (1.71–4.16)	8×10^{-6}
30	rs1525739 (C) [176]*	*LOC100287613*	7p21	EUR		NR	0.49	0.16	0.27	0.33	NR	6×10^{-6}
31	rs7636 (A) [124]	*ACHE*	7q22.1	EA		0.06	0.04	0	0.05	0.33	1.85 (1.42–2.41)	5×10^{-6}
32	rs4527850 (T) [152]	*SLA-WISP1*	8q24.22	PS		0.75	0.72	0.42	0.69	0.89	1.23 (1.13–1.34)	2×10^{-6}
33	rs564398 (T) [74]	*CDKN2B-AS1*	9p21.3	EUR		0.56	0.57	0.92	0.67	1.00	1.13(1.08–1.19)	1×10^{-6}
34	rs7020996 (C) [109]	*CDKN2B-AS1-DMRTA1*	9p21.3	EUR		NR	0.81	0.57	—	0.74	1.26(1.15–1.38)	2×10^{-7}
35	rs649891 (C) [150]	*PTPRD*	9p23	MA		0.35	0.20	0.73	0.43	0.79	NR	6×10^{-6}
36	rs10993738 (C) [138]	*SYK*	9q22.2	EA		0.15	0.02	0.26	—	0	1.16 (1.09–1.23)	5×10^{-6}
37	rs773506 (G) [121]	*SYK-AUH*	9q22.31	AA		0.77	0.63	0.25	0.50	0.13	1.32 (1.18–1.49)	6×10^{-6}
38	rs10980508 (A) [176]*	*SVEP1-RPS21P5*	9q31	EUR		NR	0.86	0.97	0.97	0.94	NR	1×10^{-6}
39	rs1327796 (G) [138]	*PALM2*	9q31.3	EA		0.24	0.22	0.27	—	0.21	1.13 (1.08–1.20)	3×10^{-6}
40	rs6583826 (G) [124]	*IDE-RPL11P4*	10q23.33	SEA (M, C, I)		0.26	0.53	0.27	0.33	0.50	1.18 (1.10–1.27)	7×10^{-6}
41	rs10741243 (G) [124]	*TCERG1L*	10q26.3	SEA (I)		0.93	0.95	0.89	—	0.49	1.75 (1.38–2.23)	5×10^{-6}
42	rs9300039 (C) [98]	*RPL9P23-HNRNPKP3*	11p12	EUR		0.89	0.87	0.70	0.82	0.85	1.48 (1.28–1.71)	6×10^{-8}

Table 2: Continued.

Number	SNP (allele)[1,2]	Mapped Gene(s)	Region[3]	Disc[4] Pop	Rep[5] Pop	RAF[6]	RAF_{EU}[7]	RAF_{EA}[8]	RAF_{SA}[9]	RAF_{AF}[10]	OR (95% CI)	*P* value
43	rs2722769 (C) [148]	*HMGN1P22-MTND5P21*	11p15.3	AA		0.53	0.56	0.56	0.76	0.99	1.35 (1.19–1.54)	2×10^{-6}
44	rs7107217 (C) [148]	*RPS27P20-TMEM45B*	11q24.3	AA		0.91	0.50	0.34	0.65	0.54	1.18 (1.10–1.27)	3×10^{-7}
45	rs12304921 (G) [104]	*HIGD1C*	12q13.12	EUR		0.15	0.16	0.50	0.36	0.15	2.5 (1.53–4.09)	7×10^{-6}
46	rs1153188 (A) [109]	*DCD-VDAC1P5*	12q13.2	EUR		0.73	0.74	0.99	0.83	0.79	1.08 (1.05–1.11)	2×10^{-7}
47	rs2358944 (G) [121]	*PCNPP3-RPSAP52*	12q14.3	AA		0.77	0.14	0.67	0.38	0.89	1.33 (1.18–1.49)	4×10^{-6}
48	rs1495377 (G) [104]	*TSPAN8-LGR5*	12q21.1	EUR		0.50	0.48	0.24	0.41	0.15	1.28 (1.14–1.49)	7×10^{-6}
49	rs4760790 (A) [116]	*TSPAN8-LGR5*	12q21.1	EUR		NR	0.22	0.24	—	0.14	1.11 (1.06–1.16)	4×10^{-6}
50	rs730570 (G) [149]	*BEGAIN-DLK1*	14q32.2	HIS		NR	0.16	0.80	0.45	0.80	1.14 (1.08–1.21)	8×10^{-6}
51	rs1436953 (G) [145]	*C2CD4A-C2CD4B*	15q22.2	EA		0.64	0.43	0.57	0.57	0.24	1.14 (NR)	8×10^{-6}
52	rs1436955 (C) [120]	*C2CD4A-C2CD4B*	15q22.2	EA		0.73	0.74	0.74	0.75	0.65	1.13 (1.08–1.19)	7×10^{-7}
53	rs7119 (T) [124]	*HMG20A*	15q24.3	SEA (M, C, I)		0.19	0.40	0.17	0.24	0.38	1.24 (1.14–1.34)	5×10^{-7}
54	rs17177078 (C) [176]*	*TNRC6A*	16p12	EUR		NR	0.93	1.00	0.97	1.00	NR	5×10^{-6}
55	rs16955379 (C) [127]	*CMIP*	16q23.2	EA		0.80	0.98	0.77	—	0.96	1.08 (1.05–1.12)	3×10^{-7}
56	rs17797882 (T) [127]	*RPS3P7-MAF*	16q23.2	EA		0.32	0.01	0.18	0.04	0.05	1.08 (1.05–1.12)	9×10^{-7}
57	rs623323 (T) [152]	*RNMTL1-NXN*	17p13.3	PS		0.15	0.20	0.11	0.13	0.50	1.28 (1.15–1.42)	4×10^{-6}
58	rs10460009 (C) [124]	*LPIN2; LOC727896*	18p11.31	SEA (M)		0.60	0.92	0.53	0.73	0.95	1.35 (1.18–1.54)	9×10^{-6}
59	rs472265 (G) [124]	*PAPL*	19q13.2	SEA (I)		0.22	0.16	0.26	0.27	0.26	1.39 (1.20–1.61)	9×10^{-6}
60	rs328506 (C) [152]	*RBM38-HMGB1P1*	20q13.31	PS	SA	0.80	0.72	1.00	0.90	0.72	1.11 (1.06–1.15)	2×10^{-6}
61	rs2833610 (A) [124]	*HUNK-MIS18A*	21q22.11	SEA (M, C, I)		0.57	0.30	0.57	0.51	0.32	1.17 (1.09–1.24)	4×10^{-6}
62	rs2106294 (T) [121]	*LIMK2*	22q12.2	AA		0.94	0.70	0.86	0.75	1.00	1.75 (1.39–2.22)	4×10^{-6}
63	rs470089 (G) [176]*	*SULT4A1*	22q13.3	EUR		NR	0.8	0.93	0.76	0.60	NR	9×10^{-6}

[1]*The* SNP-risk allele: SNP(s) most strongly associated with trait (risk allele).
[2]Reference for the largest study reporting association of the SNP with T2D or fasting plasma glucose at genome-wide significance ($P < 5 \times 10^{-8}$).
[3]Cytogenetic region associated with the SNP (NCBI).
[4]Discovery population; EUR: European; SA: South Asian; EA: East Asian; SEA: Southeast Asian; AA: African-American, MA: Mexican-American; HIS: Hispanic; PS: Punjabi Sikh; M: Malay; C: Chinese; I: Indian.
[5]Replication population: it has been confirmed in other populations; EUR: European; SA: South Asian; EA: East Asian; SEA: Southeast Asian; AA: African- American, MA: Mexican American; HIS: Hispanic; PS: Punjabi Sikh; M: Malay; C: Chinese; I: Indian.
[6]Reported risk allele frequency (RAF) for the SNP; NR if not reported.
[7]RAF in HapMap population for Utah residents with Northern and Western European-ancestry from the CEPH collection; "—" denotes data not listed in HapMap.
[8]RAF in HapMap population for Han Chinese in Beijing, China; "—" denotes data not listed in HapMap.
[9]RAF in HapMap population for Gujarati Indians in Houston, Texas; "—" denotes data not listed in HapMap.
[10]RAF in HapMap population for Yoruban in Ibadan, Nigeria; "—" denotes data not listed in HapMap.
*Risk Allele and RAF not reported, but chosen based on minor allele frequency (MAF) in the population mentioned in the original publication.

employed meta-analysis to increase sample size and hence power to detect additional loci of similar or smaller effect. The first T2D GWAS meta-analysis was published in 2008 and was also a European study [109], representing collaboration between three different consortia; the WTCCC, DGI, and the Finland—United States Investigation of NIDDM Genetics (FUSION) which combined to form the Diabetes Genetics Replication and Meta-Analysis (DIAGRAM) consortium. This study utilised an enlarged discovery sample of 4,549 cases and 5,579 controls with replication in further 24,194 cases and 55,598 controls, all of European-ancestry. This study identified associated variants at six additional novel loci: *JAZF1, CDC123, TSPAN8* and *THADA* which are associated with β-cell dysfunction [110, 111], *ADAMTS9* which is associated with insulin action [111], and *NOTCH2*, associated with glucose-stimulated glucagon secretion by pancreatic islet cells [112].

These initial T2D GWAS were all restricted to populations of European-ancestry. The first two large-scale T2D GWAS conducted in Asian populations were reported in 2008, each employing a multi-stage design in East Asian groups. Both studies reported association of variants in *KCNQ1*, encoding the alpha subunit of a voltage-gated potassium channel expressed in the pancreas [113, 114]. These studies clearly demonstrated the utility of extending T2D GWAS to non-European populations; association of the *KCNQ1* variants with T2D was not detected in previous European GWAS, due to markedly lower frequency of the risk allele in Europeans (5% versus 40%), resulting in dramatically reduced power. A European meta-analysis subsequently confirmed association of the *KCNQ1* variants with T2D in Europeans but at significance levels far below thresholds usually inspiring replication or follow-up studies ($P \sim 0.02$).

A European study published in 2009 used multiple samples of French, Danish, and Finnish ancestry to identify association of variants in the insulin receptor substrate 1 gene (*IRS1*), showing that the risk allele is also associated with insulin resistance and hyperinsulinaemia in large population-based cohorts [115]. This contrasted with the apparent biology of previous associations, which principally related to impaired pancreatic β-cell function.

This first wave of T2D GWAS was succeeded by a second wave beginning in 2010, in which existing and new datasets were combined into expanded meta-analyses. The most notable was a large European study reported by Voight and colleagues, involving ~42,000 T2D cases and 100,000 controls split between discovery and replication stages and identifying twelve new associated loci. These included X-chromosomal association and an *HNFA1A* locus overlapping with the locus implicated in Mendelian monogenic (single gene) forms of diabetes [116]. Other studies reported at this time included three East Asian studies, one African American, and one European study, which together identified nine (9) additional loci [117–121]. Three of these genes have unknown function (*RBMS1, PTPRD,* and *SRR*) [117, 118], while *RPS12, LIMK2,* and *AUH* are associated with diabetic nephropathy [121], *C2CD4A* is associated with β-cell dysfunction [119, 122], *SPRY2* is associated with obesity and insulin resistance [120, 123], and *SASH1* is associated with insulin growth factors [121].

A subsequent 2011 meta-analysis included three Southeast Asian populations: Chinese (3955 subjects), Indian (2146 subjects), and Malay (2034 subjects) and it further emphasized the value of surveying diverse ethnic groups [124]. This study was the first to include individuals from the Malay population, the largest group in Southeast Asia, with a population size of more than 300 million [124]. This study alone contributed an additional 16 novel loci, in spite of its relatively modest sample size; this partly reflected higher minor allele frequencies in Southeast Asian populations at some associated loci (e.g., rs3792615, number 18 in Table 2).

The first T2D GWAS in South Asian populations was also published in 2011, including individuals from India, Pakistan, Sri Lanka and Bangladesh. Using a relatively modest sample size (5,561 cases and 14,458 controls in the discovery step) five additional novel T2D loci were discovered [92]: HNF4A, involved in monogenic forms of diabetes and associated with β-cell development [125]; *GRB14* which is associated with obesity and insulin resistance [126]; and another three loci with less clear functions; *AP3S2*, *ST6GAL1* and *VPS26A* [92].

Another large meta-analysis in East Asian groups were performed in 2012 and identified 10 further novel loci [127] with mostly unknown function except for *GLIS3*, associated with β-cell development [128]. It is interesting that the *MAEA* variant discovered in this study is unique to East Asian and African populations, being monomorphic in Europeans and South Asians (Table 1; number 24) [129]. Several other variants identified in this study have substantially higher risk allele frequency (RAF) in East Asians than Europeans, for example, *ZFAND3* (Table 1; number 28; 34% versus 12%), *FITM2-R3HDML* (Table 1; number 111; 41% versus 18%), and *RPS3P7-MAF* (Table 2; number 56, 18% versus 1%), enhancing their detection in East Asian samples of relatively modest size.

Reflecting the success of initial T2D GWAS and the fast pace of technology, in 2012 Voight [130] and colleagues designed the "Metabochip," a custom genotyping array enriched for variants shown to be associated with cardiometabolic traits via GWAS. These traits include T2D, coronary heart disease, myocardial infarction, body mass index, glucose and insulin level, lipid levels, and blood pressure. This new platform offers a powerful and cost-effective approach to both the discovery and follow-up of variants associated with these related traits, due to comprehensive coverage of previously associated loci (~120,000 SNPs) [130]. Morris and colleagues used the Metabochip to discover and characterize T2D-associated variants via meta-analysis of 34,840 cases and 114,981 controls of European descent, finding ten novel loci [131] not reported in previous European studies, all of which reached genome-wide significance. Another study using the Metabochip combined newly available samples with samples from previous discovery meta-analyses, using genotype data for 66,000 follow-up SNPs. This study identified 41 novel glycaemic associations, 33 of which were also associated with T2D [132]. This study implicated new loci in the aetiology of T2D and increased the overlap between loci associated with both glycaemia and T2D. These studies

highlight the Metabochip as a promising tool for identifying novel and robustly associated loci, facilitating future research into underlying biology.

Taken together, the results of T2D GWAS signify tremendous progress in our quest to understand the genetic causes of T2D. Alternatively, they also highlight the genetic complexity of this disease. Genetic variants showing replicable association with T2D uniformly exert only a modest effect on disease risk, with per-allele odds ratios typically in the range of 1.1–1.3 (Table 1). The combined effect of all variants reported to date explains only about 10% of observed familial clustering [116]. Furthermore, the functional significance of various loci remains unclear. While some appear to be associated with β-cell function and insulin resistance, the biological role of many of them remains unknown. This suggests that the findings to date represent the first stage of a long journey to understanding T2D genetic risk.

2.7. Polygenic Models of T2D. The distribution of odds ratios observed for validated T2D-associated SNPs suggests that numerous, associated loci exist with even smaller effects than those identified to date. One would not expect such loci to have reached genome-wide significance in previous GWAS due to insufficient power. The existence of such additional small effect loci is consistent with the pattern of additional associated variants being discovered as sample sizes have increased; it is also consistent with validated SNP associations explaining only a small proportion of the T2D heritability estimated from twin studies, known as the "missing heritability" problem.

Two methods have recently been developed for assessing the contribution of common SNPs not reaching genome-wide significance to the heritability of a trait. These are polygenic scoring [133] and mixed linear modelling [134]. Both methods test the combined effects of thousands (or hundreds of thousands) of SNPs upon a trait of interest. A recent study by Stahl and colleagues used polygenic analyses and linear mixed modelling to show that thousands of SNPs contribute to T2D risk, estimating that about 50% of observed variance in T2D risk could be attributed to the combined effects of all SNPs genome-wide [135]. These investigators suggested that at least 70% of T2D heritability can be attributed to common SNPs represented on GWAS arrays [135], with most having very small individual effects upon disease risk.

2.8. Population Differences in T2D Risk Alleles. The frequency of T2D risk alleles often varies between populations, producing population differences in the attributable risk due to a particular genetic risk factor or combination of risk factors. The discovery of *KCNQ1* emphasized the impact of such frequency differences upon genetic discoveries [136]. Association of *KCNQ1* variants was found in East Asian populations [113, 114] using a considerably smaller sample size than that required to detect the association in with European populations [116], reflecting higher allele frequency (33% in East Asian versus 8% in Europeans) and hence statistical power in Asian groups. In addition, variants at the *TCF7L2* locus showed the inverse; a high risk allele frequency in Europeans (30%) compared to a low frequency in East Asians (3%) enhanced the detection in European studies [137]. Similarly, the *SYK* variant demonstrates a RAF of 26% in East Asians [138]and only 2% in Europeans and is monomorphic in Africans (Table 2; number 36). Further, a number of T2D risk variants are monomorphic (not variable) in some populations, preventing the detection of an association in these groups. The recently reported *SCGG* variant is unique to Indian Punjabi Sikh, being monomorphic in both European and African populations (Table 1; number 92). Other instances include the *THADA* variant, which was discovered in European populations but is monomorphic in East Asians (Table 1; number 4), *RND3-FABP5P10* that was discovered in African Americans but is monomorphic in Europeans (Table 1; number 6), and *G6PC2*, discovered in Europeans but monomorphic in Africans (Table 1; number 10). For a set of SNPs showing association with T2D across multiple populations, Table 3 shows risk allele frequencies and odds ratios for different populations in which associations have been reported. For these 14 SNPs, risk allele frequencies commonly differ across populations; however, allelic effects upon disease seem markedly consistent in both direction and magnitude, given overlapping confidence intervals for allelic odds ratios. Taken together, these results suggest that population differences can have important effects on power to *detect* common genetic associations for T2D in samples of diverse ancestry but may have less impact upon disease risk within individuals carrying the identified risk alleles. Nevertheless, at the population level, the attributable risk of such genetic variants will increase with allele frequency, thus potentially influencing population disease burden.

Significantly, a recent study assessing thousands of genetic associations showed that T2D has the most extreme population differentiation of risk allele frequencies among a broad range of complex diseases [139]. T2D risk allele frequencies demonstrated clear gradient matching paths of early human migration, suggesting potential differences in evolutionary adaptation to food availability, dietary patterns, or agricultural practices. This is consistent with "thrifty genotype" hypothesis [139, 140], which proposes that susceptibility to obesity and T2D in some populations reflects historical positive selection for genotypes promoting efficiency of metabolism, and energy and fat storage, thus providing an advantage in times of nutrient shortage [141]. This might explain the extraordinarily high prevalence of diabetes now seen among certain populations [34, 142, 143], potentially reflecting historical feast and famine cycles [62], increasing the frequency of thrifty genotypes and genetic predisposition to obesity and diabetes. While being unproven, this may partly explain higher susceptibility to abdominal obesity at lower BMI and reduced muscle mass with increased insulin resistance in Asian compared with Caucasian populations [7]. Nevertheless, pronounced population differentiation of T2D risk allele frequencies provides a strong rationale for further comprehensive genetic studies of T2D in diverse populations, expanding on the comprehensive studies in European samples.

To date, a range of non-European T2D GWAS have been conducted, including studies in Japanese [114, 119, 138, 144],

Table 3: Population-specific odds ratios and risk allele frequencies for SNPs associated with T2D in multiple populations.

Number	SNPs	Ref allele	Disc Pop	RAF	OR (95% CI)	Ref allele	1st Rep Pop	RAF	OR (95% CI)	Ref allele	2nd Rep Pop	RAF	OR (95% CI)
1	rs1470579	C	EUR [116]	0.3	1.14 (1.09–1.19)	C	EA [177]	0.24	1.33 (1.20–1.48)	C	SA [152]	0.5	1.06 (1.04–1.08)
2	rs4402960	T	EUR [98]	0.3	1.14 (1.11–1.18)	T	EA [144]	0.22	1.14 (1.08–1.21)				
3	rs7754840	C	EUR [103]	0.34	1.12 (1.08–1.16)	C	EA [146]	0.4	1.35 (1.23–1.48)				
4	rs13266634	C	EUR [74]	0.76	1.12 (1.07–1.16)	C	EA [144]	0.53	1.22 (1.15–1.28)				
5	rs2383208	A	EA [144]	0.59	1.34 (1.27–1.41)	A	SA [151]	0.91	1.23 (1.13–1.34)				
6	rs1118875	C	EUR [98]	0.58	1.13 (1.09–1.17)	C	EA [144]	0.34	1.21 (1.15–1.28)				
7	rs5015480	C	EUR [74]	0.58	1.13 (1.07–1.19)	C	EA [120]	0.21	1.17 (1.11–1.24)				
8	rs7903146	T	EUR [99]	0.28	1.65 (1.28–2.02)	T	EA [144]	0.03	1.54 (1.36–1.74)	T	SA [178]	0.28	1.33 (1.19–1.49)
9	rs2237892	C	EA [113]	0.67	1.40 (1.34–1.47)	C	EA [113]	0.93	1.29 (1.11–1.50)	C	HIS [149]	0.79	1.09 (1.06–1.12)
10	rs7178572	G	SA [92]	0.44	1.09 (1.06–1.12)	G	EUR [171]	0.68	1.11(1.07–1.15)				
11	rs8050136	A	EUR [98]	0.46	1.17 (1.12–1.22)	A	SA [151]	0.25	1.16 (1.09–1.24)				
12	rs4430796	G	EUR [116]	0.51	1.14 (1.08–1.20)	G	EA [146]	0.25	1.19 (1.13–1.25)				
13	rs5945326	A	EUR [116]	0.78	1.27 (1.18–1.37)	A	EA [146]	0.66	1.18 (1.13–1.23)				
14	rs4712523	G	EA [144]	0.41	1.27 (1.21–1.33)	G	EUR [115]	0.34	1.20 (1.14–1.26)				

Population. EUR: European; SA: South Asian; EA: East Asian; HIS: Hispanic.

Chinese [117, 145, 146], African-American [121, 147, 148], Southeast Asian [124], Hispanic [149], Mexican-American [150], South Asian [92], Indo-European [151], and Indian Punjabi Sikh [152]. These studies have led to new discoveries, including novel loci and loci that seem specific to certain populations [119, 127, 151, 152]. While many loci appear to contribute broadly to T2D risk, some loci have currently been confirmed in European populations only, including *WFS1, NOTCH2, THADA, ADAMTS9, TSPAN8/LGR5, INS-IGF2, ADCY5, GCK, MTRNR1B, HMGA2, HNF1A, ZBED3, KLF14, ZFAND6, PRC1, TLES/CHCHD9,* and *RBMS1* [109, 116, 153–155]. Other loci currently show association specifically in East Asian populations, including *PTPRD, SRR, CDC123/CAMK1D, PSMD6, MAEA, ZFAND3, KCNK16, GCC1/PAX4, GLIS3,* and *PEPD* [117, 119, 120, 127]. On the other hand, *TMEM163* [151] and *SGCG* [152] appear unique to South Asian and Indian Punjabi Sikh, respectively. Some of these discoveries may reflect the impact of population allele frequency differences, as previously discussed. In such cases, larger studies may eventually show that some loci contribute to disease across a broader range of populations.

Seemingly population-specific genetic associations for T2D may also reflect differences in the patterns of genomic correlation, or linkage disequilibrium (LD), between associated marker loci and the underlying unobserved functional variants. Populations with different demographic histories will often display different patterns of LD reflecting population differences in evolutionary recombination [156]. Older populations such as those in Africa have lower LD and can be helpful for finely localizing a risk variant following an initial association finding. This is because the genomic distance between disease-associated markers and true risk variants is likely to be smaller in such populations [157].

Thus, the apparent population-specificity of some known T2D risk alleles may reflect population differences in risk allele frequencies or LD between tagging and causal variants, rather than actual population-specificity of the underlying functional risk loci. We note that population-specific estimates of disease variance explained by all known T2D loci—although not widely reported—do seem largely similar between European and Asian populations. In their large 2012 study, Morris and colleagues [131] estimated that known common variants explain about 5.7% of T2D disease variance in Europeans. In 2013, Tabassum and colleagues [151] estimated that known loci combined with one novel Indian-specific locus explained 7.65% of T2D risk variance in South Asian Indians. The slightly higher estimate in Indians may potentially be explained by the inclusion of additional variants discovered between publications of the two studies, together with the inclusion of the Indian-specific locus discovered in the Tabassum study. Thus, available evidence thus does not strongly suggest that differences in the cumulative variance explained by known common T2D risk alleles can explain the markedly higher T2D prevalence observed in South Asians.

Interestingly, however, very recent studies show that population differences in linkage disequilibrium (LD) and the presence of multiple independent variants within a locus can markedly influence estimates of variance explained by known risk variants [158, 159]. Detailed fine mapping of T2D susceptibility loci in diverse populations, combined with the identification of underlying functional variants, may thus reveal population differences in the contribution of known loci to disease. Future research may also show the extent to which population differences in T2D risk can be explained by rare alleles, gene-environment interactions, or epigenetic effects.

2.9. Gene-Environment Interactions in T2D. In addition to the effects of specific genetic and environmental risk factors, gene-environment interactions are likely important mediators of population differences in T2D risk and contributors to the "missing heritability" problem. Indeed, given the relative stability of DNA sequence within populations over decades, recent increases in T2D prevalence must largely reflect environmental changes. Accordingly, the single largest contributor to T2D risk is obesity, and the global T2D epidemic chronologically parallels the global obesity epidemic.

A paucity of studies has examined gene-environment interactions in the context of T2D in general, let alone in Asian populations. A study by Qi and colleagues [160] found that a high genetic risk score formed from 10 T2D-associated SNPs was further increased by the presence of a "Westernized" dietary pattern characterised by increased red and processed meat intake and reduced dietary fibre [160]. The Westernized diet was not associated with increased risk among those with a low genetic risk score. Several studies have also found evidence for interactions between T2D-associated variants in *TCF7L2* and the quality and quantity of ingested carbohydrates in the context of T2D risk [161–163]. These studies support a possible contribution of gene-environment interactions to T2D risk, together with a potential model where interactions between recent lifestyle transitions and genetic risk factors may be contributing to the rapidly increasing prevalence of T2D in Asian populations. However, these preliminary findings require validation. Future analyses in well-designed, well-powered studies will help to clarify the role of gene-environment interactions in population differences in T2D risk.

2.10. Epigenetics. Similar to the "thrifty genotype" hypothesis, the "thrifty phenotype" hypothesis considers the adaptive consequences of the environment *in utero*. The hypothesis relates to the metabolic consequences of fetal malnutrition, proposing that adaptation to a low-calorie intrauterine environment induces permanent changes in chromatin structure and gene expression that influence insulin secretion and resistance, promoting more efficient energy utilisation and thus fetal survival [164]. According to the hypothesis, such epigenetic changes may predispose to insulin dysregulation, obesity, and T2D in later life. In support, epidemiologic studies have shown associations between small birth size, a marker for fetal malnutrition, and adult-onset T2D [165, 166]. A study by van Hoek and colleagues [167] in the Dutch Famine Birth Cohort detected an interaction between an *IGF2BP2* polymorphism and prenatal famine upon glucose level in the offspring. Interactions between other T2D risk

variant alleles and birthweight have also been associated with increased T2D risk [168, 169].

3. Conclusions

We have discussed differences in prevalence, risk factor profiles, and genetic risk allele frequencies between different Asian countries and between Asian and other continental populations. Given these differences, continued T2D genetic studies in diverse populations are likely to contribute crucially to the broadening terrain of shared and unique population genetic effects for this disorder. Future studies will ideally include large, population-specific characterisation of risk variants, studies of gene-environment interaction, and epigenetic studies. Well-powered, well-designed studies performed in diverse Asian populations should enhance the benefits of genetic discoveries and their ultimate clinical translation for these large susceptible groups.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Noraidatulakma Abdullah is sponsored by Ministry of Higher Education of Malaysia and Universiti Kebangsaan Malaysia. Elizabeth Holliday is funded by the NHMRC fellowship scheme.

References

[1] International Diabetes Federation, *IDF Diabetes Atlas*, 2012, http://www.idf.org/diabetesatlas.

[2] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes: estimates for the year 2000 and projections for 2030," *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004.

[3] R. Sicree, J. Shaw, and P. Zimmet, "Prevalence and projections," in *Diabetes Atlas*, D. Gan, Ed., pp. 16–104, International Diabetes Federation, Brussels, Belgium, 2006.

[4] T. Scully, "Diabetes in numbers," *Nature*, vol. 485, no. 7398, pp. S2–S3, 2012.

[5] C. J. Murray, T. Vos, R. Lozano et al., "Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2197–2223, 2012.

[6] U. N. H. L. A. B. I. UnitedHealth, National Heart, Lung, and Blood Institute Centers of Excellence, M. T. Cerqueira, A. Cravioto et al., "Global response to non-communicable disease," *BMJ (Clinical research ed.)*, vol. 342, p. d3823, 2011.

[7] J. C. N. Chan, V. Malik, W. Jia et al., "Diabetes in Asia: epidemiology, risk factors, and pathophysiology," *Journal of the American Medical Association*, vol. 301, no. 20, pp. 2129–2140, 2009.

[8] S. Nakanishi, M. Okubo, M. Yoneda, K. Jitsuiki, K. Yamane, and N. Kohno, "A comparison between Japanese-Americans living in Hawaii and Los Angeles and native Japanese: the impact of lifestyle westernization on diabetes mellitus," *Biomedicine and Pharmacotherapy*, vol. 58, no. 10, pp. 571–577, 2004.

[9] R. Oza-Frank, M. K. Ali, V. Vaccarino, and K. M. V. Narayan, "Asian Americans: diabetes prevalence across U.S. and World Health Organization weight classifications," *Diabetes Care*, vol. 32, no. 9, pp. 1644–1646, 2009.

[10] J. Ye, G. Rust, P. Baltrus, and E. Daniels, "Cardiovascular risk factors among Asian Americans: results from a national health survey," *Annals of Epidemiology*, vol. 19, no. 10, pp. 718–723, 2009.

[11] V. Bhalla, C. W. Fong, S. K. Chew, and K. Satku, "Changes in the levels of major cardiovascular risk factors in the multiethnic population in Singapore after 12 years of a national noncommunicable disease intervention programme," *Singapore Medical Journal*, vol. 47, no. 10, pp. 841–850, 2006.

[12] Institute of Public Health, *National Health Morbidity Survey III*, The Ministry of Health Malaysia, Kuala Lumpur, Malaysia, 2006.

[13] Institute for Public Health, *National Health and Morbidity Survey*, 2011.

[14] K. G. M. M. Alberti, "The classification and diagnosis of diabetes mellitus," in *Textbook of Diabetes*, C. S. Cockram, R. I. G. Holt, A. Flyvbjerg, and B. J. Goldstein, Eds., pp. 24–30, Blackwell Publishing Company, London, UK, 2010.

[15] "World Health OrganizationDiabetes Fact Sheet," 2013, http://www.who.int/mediacentre/factsheets/fs312/en/index.html.

[16] D. R. Matthews, J. P. Hosker, and A. S. Rudenski, "Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man," *Diabetologia*, vol. 28, no. 7, pp. 412–419, 1985.

[17] A. Raji, E. W. Seely, R. A. Arky, and D. C. Simonson, "Body fat distribution and insulin resistance in healthy Asian Indians and Caucasians," *Journal of Clinical Endocrinology and Metabolism*, vol. 86, no. 11,pp. 5366–5371, 2001.

[18] A. K. Manning, M. F. Hivert, R. A. Scott et al., "A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance," *Nature Genetics*, vol. 44, no. 6, pp. 659–669, 2012.

[19] H. Gao, A. Salim, J. Lee, E. S. Tai, and R. M. van Dam, "Can body fat distribution, adiponectin levels and inflammation explain differences in insulin resistance between ethnic Chinese, Malays and Asian Indians?" *International Journal of Obesity*, vol. 36, no. 8, pp. 1086–1093, 2012.

[20] P. M. McKeigue, M. G. Marmot, Y. D. Syndercombe Court, D. E. Cottier, S. Rahman, and R. A. Riemersma, "Diabetes, hyperinsulinaemia, and coronary risk factors in Bangladeshis in East London," *British Heart Journal*, vol. 60, no. 5, pp. 390–396, 1988.

[21] P. M. McKeigue, B. Shah, and M. G. Marmot, "Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians," *The Lancet*, vol. 337, no. 8738, pp. 382–386, 1991.

[22] J. Dhawan, C. L. Bray, R. Warburton, D. S. Ghambhir, and J. Morris, "Insulin resistance, high prevalence of diabetes, and cardiovascular risk in immigrant Asians," *British Heart Journal*, vol. 72, no. 5, pp. 413–421, 1994.

[23] S. Dickinson, S. Colagiuri, E. Faramus, P. Petocz, and J. C. Brand-Miller, "Postprandial hyperglycemia and insulin sensitivity differ among lean young adults of different ethnicities," *Journal of Nutrition*, vol. 132, no. 9, pp. 2574–2579, 2002.

[24] C.-F. Liew, E.-S. Seah, K.-P. Yeo, K.-O. Lee, and S. D. Wise, "Lean, nondiabetic Asian Indians have decreased insulin sensitivity and insulin clearance, and raised leptin compared to Caucasians and Chinese subjects," *International Journal of Obesity*, vol. 27, no. 7, pp. 784–789, 2003.

[25] S. E. Kahn, "The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes," *Diabetologia*, vol. 46, no. 1, pp. 3–19, 2003.

[26] B. Ahrén and H. Larsson, "Quantification of insulin secretion in relation to insulin sensitivity in nondiabetic postmenopausal women," *Diabetes*, vol. 51, supplement 1, pp. S202–S211, 2002.

[27] M. H. Black, R. M. Watanabe, E. Trigo et al., "High-fat diet is associated with obesity-mediated insulin resistance and beta-cell dysfunction in Mexican Americans," *Journal of Nutrition*, vol. 143, no. 4, pp. 479–485, 2013.

[28] K. Ohtsubo, M. Z. Chen, J. M. Olefsky, and J. D. Marth, "Pathway to diabetes through attenuation of pancreatic beta cell glycosylation and glucose transport," *Nature Medicine*, vol. 17, no. 9, pp. 1067–1076, 2011.

[29] M. Fukushima, M. Usami, M. Ikeda et al., "Insulin secretion and insulin sensitivity at different stages of glucose tolerance: a cross-sectional study of Japanese type 2 diabetes," *Metabolism*, vol. 53, no. 7, pp. 831–835, 2004.

[30] D. Tripathy, M. Carlsson, P. Almgren et al., "Insulin secretion and insulin sensitivity in relation to glucose tolerance: lessons from the Botnia Study," *Diabetes*, vol. 49, no. 6, pp. 975–980, 2000.

[31] K. Mørkrid, A. K. Jenum, L. Sletner et al., "Failure to increase insulin secretory capacity during pregnancy-induced insulin resistance is associated with ethnicity and gestational diabetes," *European Journal of Endocrinology*, vol. 167, no. 4, pp. 579–588, 2012.

[32] D. W. Bowden, A. J. Cox, B. I. Freedman et al., "Review of the Diabetes Heart Study (DHS) family of studies: a comprehensively examined sample for genetic and epidemiological studies of type 2 diabetes and its complications," *The Review of Diabetic Studies*, vol. 7, no. 3, pp. 188–201, 2010.

[33] N. J. Morrish, S.-L. Wang, L. K. Stevens et al., "Mortality and causes of death in the WHO multinational study of vascular disease in diabetes," *Diabetologia*, vol. 44, supplement 2, pp. S14–S21, 2001.

[34] K. K. Yeo, B. C. Tai, D. Heng et al., "Ethnicity modifies the association between diabetes mellitus and ischaemic heart disease in Chinese, Malays and Asian Indians living in Singapore," *Diabetologia*, vol. 49, no. 12, pp. 2866–2873, 2006.

[35] K. A. Earle, K. K. Porter, J. Ostberg, and J. S. Yudkin, "Variation in the progression of diabetic nephropathy according to racial origin," *Nephrology Dialysis Transplantation*, vol. 16, no. 2, pp. 286–290, 2001.

[36] B. A. Young, C. Maynard, and E. J. Boyko, "Racial differences in diabetic nephropathy, cardiovascular disease, and mortality in a national population of veterans," *Diabetes Care*, vol. 26, no. 8, pp. 2392–2399, 2003.

[37] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 614–618, 2012.

[38] S. Sivaprasad, B. Gupta, M. C. Gulliford et al., "Ethnic variations in the prevalence of diabetic retinopathy in people with diabetes attending screening in the United Kingdom (DRIVE UK)," *PLoS ONE*, vol. 7, no. 3, Article ID e32182, 2012.

[39] M. K. Ali, M. B. Weber, and K. M. V. Narayan, "The global burden of diabetes, in Textbook of Diabetes," C. S. Cockram, R. I. G. Holt, A. Flyvbjerg, and B. J. Goldstein, Eds., pp. 69–84, Balckwell Publishing, USA, 2010.

[40] G. Roglic, N. Unwin, P. H. Bennett et al., "The burden of mortality attributable to diabetes: realistic estimates for the year 2000," *Diabetes Care*, vol. 28, no. 9, pp. 2130–2135, 2005.

[41] "World Health Organization Obesity and overweight Fact-sheet," 2013.

[42] R. L. Westley and F. E. May, "A twenty-first century cancer epidemic caused by obesity: the involvement of insulin, diabetes, and insulin-like growth factors," *International Journal of Endocrinology*, vol. 2013, Article ID 632461, 37 pages, 2013.

[43] M. B. Weber, R. Oza-Frank, L. R. Staimez, M. K. Ali, and K. M. Narayan, "Type 2 diabetes in Asians: prevalence, risk factors, and effectiveness of behavioral intervention at individual and population levels," *Annual Review of Nutrition*, vol. 32, pp. 417–439, 2012.

[44] P. Boffetta, D. McLerran, Y. Chen et al., "Body mass index and diabetes in Asia: a cross-sectional pooled analysis of 900,000 individuals in the Asia cohort consortium," *PLoS ONE*, vol. 6, no. 6, Article ID e19930, 2011.

[45] P. Deurenberg, M. Deurenberg-Yap, and S. Guricci, "Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship," *Obesity Reviews*, vol. 3, no. 3, pp. 141–146, 2002.

[46] S. Gurrici, Y. Hartriyanti, J. G. A. J. Hautvast, and P. Deurenberg, "Relationship between body fat and body mass index: differences between Indonesians and Dutch Caucasians," *European Journal of Clinical Nutrition*, vol. 52, no. 11, pp. 779–783, 1998.

[47] M. Deurenberg-Yap, G. Schmidt, W. A. van Staveren, J. G. A. J. Hautvast, and P. Deurenberg, "Body fat measurement among Singaporean Chinese, Malays and Indians: a comparative study using a four-compartment model and different two-compartment models," *British Journal of Nutrition*, vol. 85, no. 4, pp. 491–498, 2001.

[48] C.-H. Cheng, C.-C. Ho, C.-F. Yang, Y.-C. Huang, C.-H. Lai, and Y.-P. Liaw, "Waist-to-hip ratio is a better anthropometric index than body mass index for predicting the risk of type 2 diabetes in Taiwanese population," *Nutrition Research*, vol. 30, no. 9, pp. 585–593, 2010.

[49] Z. Xin, C. Liu, W. Y. Niu et al., "Identifying obesity indicators which best correlate with type 2 diabetes in a Chinese population," *BMC Public Health*, vol. 12, p. 732, 2012.

[50] P. Bjorntorp, "'Portal' adipose tissue as a generator of risk factors for cardiovascular disease and diabetes," *Arteriosclerosis*, vol. 10, no. 4, pp. 493–496, 1990.

[51] P. Bjorntorp, "Metabolic implications of body fat distribution," *Diabetes Care*, vol. 14, no. 12, pp. 1132–1143, 1991.

[52] J.-P. Després and I. Lemieux, "Abdominal obesity and metabolic syndrome," *Nature*, vol. 444, no. 7121, pp. 881–887, 2006.

[53] J. J. Díez and P. Iglesias, "The role of the novel adipocyte-derived hormone adiponectin in human disease," *European Journal of Endocrinology*, vol. 148, no. 3, pp. 293–300, 2003.

[54] R. I. G. Holt and C. W. Ronald, "Epidemiology of type 2 diabetes," in *Textbook of Diabetes*, R. I. G. Holt, C. S. Cockram, A. Flyvbjerg, and B. J. Goldstein, Eds., pp. 45–68, Balckwell Publishing, Hong Kong, China, 2010.

[55] S. A. Lear, K. H. Humphries, S. Kohli, A. Chockalingam, J. J. Frohlich, and C. L. Birmingham, "Visceral adipose tissue accumulation differs according to ethnic background: results of the Multicultural Community Health Assessment Trial (M-CHAT)," *American Journal of Clinical Nutrition*, vol. 86, no. 2, pp. 353–359, 2007.

[56] S. K. Kumanyika, "Obesity in minority populations: an epidemiologic assessment," *Obesity research*, vol. 2, no. 2, pp. 166–182, 1994.

[57] P. L. Lutsey, M. A. Pereira, A. G. Bertoni, N. R. Kandula, and D. R. Jacobs Jr., "Interactions between race/ethnicity and anthropometry in risk of incident diabetes," *American Journal of Epidemiology*, vol. 172, no. 2, pp. 197–204, 2010.

[58] S. A. Lear, K. H. Humphries, S. Kohli, and C. L. Birmingham, "The use of BMI and waist circumference as surrogates of body fat differs by ethnicity," *Obesity*, vol. 15, no. 11, pp. 2817–2824, 2007.

[59] S. Patel, N. Unwin, R. Bhopal et al., "A comparison of proxy measures of abdominal obesity in Chinese, European and South Asian adults," *Diabetic Medicine*, vol. 16, no. 10, pp. 853–860, 1999.

[60] A. Misra, R. Sharma, S. Gulati et al., "Consensus dietary guidelines for healthy living and prevention of obesity, the metabolic syndrome, diabetes, and related disorders in Asian Indians," *Diabetes Technology and Therapeutics*, vol. 13, no. 6, pp. 683–694, 2011.

[61] Y. Wang, J. Mi, X.-Y. Shan, Q. J. Wang, and K.-Y. Ge, "Is China facing an obesity epidemic and the consequences? The trends in obesity and chronic disease in China," *International Journal of Obesity*, vol. 31, no. 1, pp. 177–188, 2007.

[62] F. B. Hu, "Globalization of diabetes: the role of diet, lifestyle, and genes," *Diabetes Care*, vol. 34, no. 6, pp. 1249–1257, 2011.

[63] A. Ramachandran, C. Snehalatha, A. D. S. Baskar et al., "Temporal changes in prevalence of diabetes and impaired glucose tolerance associated with lifestyle transition occurring in the rural population in India," *Diabetologia*, vol. 47, no. 5, pp. 860–865, 2004.

[64] S. W. Ng, E. C. Norton, and B. M. Popkin, "Why have physical activity levels declined among Chinese adults? Findings from the 1991–2006 China health and nutrition surveys," *Social Science and Medicine*, vol. 68, no. 7, pp. 1305–1314, 2009.

[65] E. S. Ford, C. Li, and N. Sattar, "Metabolic syndrome and incident diabetes," *Diabetes Care*, vol. 31, no. 9, pp. 1898–1904, 2008.

[66] S. Rampal, S. Mahadeva, E. Guallar et al., "Ethnic differences in the prevalence of metabolic syndrome: results from a multi-ethnic population-based survey in Malaysia," *PLoS ONE*, vol. 7, no. 9, Article ID e46365, 2012.

[67] N. T. Ayas, D. P. White, W. K. Al-Delaimy et al., "A prospective study of self-reported sleep duration and incident diabetes in women," *Diabetes Care*, vol. 26, no. 2, pp. 380–384, 2003.

[68] E. G. Holliday, C. A. Magee, L. Kritharides, E. Banks, and J. Attia, "Short sleep duration is associated with risk of future diabetes but not cardiovascular disease: a prospective study and meta-analysis," *PLoS ONE*, vol. 8, no. 11,Article ID e82305, 2013.

[69] J. L. Broussard, D. A. Ehrmann, E. van Cauter, E. Tasali, and M. J. Brady, "Impaired insulin signaling in human adipocytes after experimental sleep restriction: a randomized, crossover study," *Annals of Internal Medicine*, vol. 157, no. 8, pp. 549–557, 2012.

[70] F. Landi, G. Onder, and R. Bernabei, "Sarcopenia and diabetes: two sides of the same coin," *Journal of the American Medical Directors Association*, vol. 14, no. 8, pp. 540–541, 2013.

[71] R. J. Manders, P. Jonathan, S. C. Forbes, and D. G. Candow, "Insulinotropic and muscle protein synthetic effects of branched-chain amino acids: potential therapy for type 2 diabetes and sarcopenia," *Nutrients*, vol. 4, no. 11, pp. 1664–1678, 2012.

[72] C. V. Calkin, D. M. Gardner, T. Ransom, and M. Alda, "The relationship between bipolar disorder and type 2 diabetes: more than just co-morbid disorders," *Annals of Medicine*, vol. 45, no. 2, pp. 171–181, 2013.

[73] P. R. Blackett and D. K. Sanghera, "Genetic determinants of cardiometabolic risk: a proposed model for phenotype association and interaction," *Journal of Clinical Lipidology*, vol. 7, no. 1, pp. 65–81,2013.

[74] E. Zeggini, M. N. Weedon, C. M. Lindgren et al., "Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes ," *Science*, vol. 316, no. 5829, pp. 1336–1341,2007.

[75] A. Doria, M.-E. Patti, and C. R. Kahn, "The emerging genetic architecture of type 2 diabetes," *Cell Metabolism*, vol. 8, no. 3, pp. 186–200, 2008.

[76] J. Kobberling and H. Tillil, "Empirical risk figures for first-degree relatives of non-insulin dependent diabetics," in *The Genetics of Diabetes Mellitus*, J. Kobberling and R. Tattersall, Eds., pp. 201–209, Academic Press, London, UK, 1982.

[77] B. Newman, J. V. Selby, M.-C. King, C. Slemenda, R. Fabsitz, and G. D. Friedman, "Concordance for Type 2 (non-insulin-dependent) diabetes mellitus in male twins," *Diabetologia*, vol. 30, no. 10, pp. 763–768, 1987.

[78] J. Kaprio, J. Tuomilehto, M. Koskenvuo et al., "Concordance for Type 1 (insulin-dependent) and Type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland," *Diabetologia*, vol. 35, no. 11, pp. 1060–1067, 1992.

[79] F. Medici, M. Hawa, A. Ianari, D. A. Pyke, and R. D. G. Leslie, "Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis," *Diabetologia*, vol. 42, no. 2, pp. 146–150, 1999.

[80] P. Poulsen, K. Ohm Kyvik, A. Vaag, and H. Beck-Nielsen, "Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance—a population-based twin study," *Diabetologia*, vol. 42, no. 2, pp. 139–145, 1999.

[81] S. Carlsson, A. Ahlbom, P. Lichtenstein, and T. Andersson, "Shared genetic influence of BMI, physical activity and type 2 diabetes: a twin study," *Diabetologia*, vol. 56, no. 5, pp. 1031–1035, 2013.

[82] P. Poulsen, K. Levin, I. Petersen, K. Christensen, H. Beck-Nielsen, and A. Vaag, "Heritability of insulin secretion, peripheral and hepatic insulin action, and intracellular glucose partitioning in young and old Danish twins," *Diabetes*, vol. 54, no. 1, pp. 275–283, 2005.

[83] S. N. Wulan, K. R. Westerterp, and G. Plasqui, "Ethnic differences in body composition and the associated metabolic profile: a comparative study between Asians and Caucasians," *Maturitas*, vol. 65, no. 4, pp. 315–319, 2010.

[84] M. F. P. Vaxillare, "The genetics of type 2 diabetes: from candidate gene biology to genome-wide studies," in *Textbook of Diabetes*, C. S. Cockram, R. I. G. Holt, A. Flyvbjerg, and B. J. Goldstein, Eds., pp. 191–214, Balckwell Publishing, 2010.

[85] S. Bevan and H. S. Markus, "Genetics of common polygenic ischaemic stroke: current understanding and future challenges," *Stroke Research and Treatment*, vol. 2011,Article ID 179061, 6 pages, 2011.

[86] N. Risch and K. Merikangas, "The future of genetic studies of complex human diseases," *Science*, vol. 273, no. 5281, pp. 1516–1517, 1996.

[87] K. S. Park, "The search for genetic risk factors of type 2 diabetes mellitus," *Diabetes & Metabolism*, vol. 35, no. 1, pp. 12–22, 2011.

[88] Y. Horikawa, N. Oda, N. J. Cox et al., "Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus," *Nature Genetics*, vol. 26, no. 2, pp. 163–175, 2000.

[89] D. Meyre, N. Bouatia-Naji, A. Tounian et al., "Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes," *Nature Genetics*, vol. 37, no. 8, pp. 863–867, 2005.

[90] K. Silander, K. L. Mohlke, L. J. Scott et al., "Genetic variation near the hepatocyte nuclear factor-4α gene predicts susceptibility to type 2 diabetes," *Diabetes*, vol. 53, no. 4, pp. 1141–1149, 2004.

[91] F. Vasseur, N. Helbecque, C. Dina et al., "Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians," *Human Molecular Genetics*, vol. 11,no. 21,pp. 2607–2614, 2002.

[92] J. S. Kooner, D. Saleheen, X. Sim et al., "Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci," *Nature Genetics*, vol. 43, no. 10, pp. 984–989, 2011.

[93] J. M. Kwon and A. M. Goate, "The candidate gene approach," *Alcohol Research and Health*, vol. 24, no. 3, pp. 164–168, 2000.

[94] S. Bevan, M. Traylor, P. Adib-Samii et al., "Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations," *Stroke*, vol. 43, no. 12, pp. 3161–3167, 2012.

[95] D. Altshuler, J. N. Hirschhorn, M. Klannemark et al., "The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes," *Nature Genetics*, vol. 26, no. 1, pp. 76–80, 2000.

[96] A. L. Gloyn, M. N. Weedon, K. R. Owen et al., "Large-scale association studies of variants in genes encoding the pancreatic β-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11E23K variant is associated with type 2 diabetes," *Diabetes*, vol. 52, no. 2, pp. 568–572, 2003.

[97] S. F. A. Grant, G. Thorleifsson, I. Reynisdottir et al., "Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes," *Nature Genetics*, vol. 38, no. 3, pp. 320–323, 2006.

[98] L. J. Scott, K. L. Mohlke, L. L. Bonnycastle et al., "A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants," *Science*, vol. 316, no. 5829, pp. 1341–1345, 2007.

[99] R. Sladek, G. Rocheleau, J. Rung et al., "A genome-wide association study identifies novel risk loci for type 2 diabetes," *Nature*, vol. 445, no. 7130, pp. 881–885, 2007.

[100] I. Pe'er, R. Yelensky, D. Altshuler, and M. J. Daly, "Estimation of the multiple testing burden for genomewide association studies of nearly all common variants," *Genetic Epidemiology*, vol. 32, no. 4, pp. 381–385, 2008.

[101] Q. Qi and F. B. Hu, "Genetics of type 2 diabetes in European populations," *Journal of Diabetes*, vol. 4, no. 3, pp. 203–212, 2012.

[102] F. Chimienti, S. Devergnas, A. Favier, and M. Seve, "Identification and cloning of a β-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules," *Diabetes*, vol. 53, no. 9, pp. 2330–2337, 2004.

[103] Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels," *Science*, vol. 316, no. 5829, pp. 1331–1336, 2007.

[104] Wellcome Trust Case Control Consortium, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, no. 7145, pp. 661–678, 2007.

[105] M. J. Groenewoud, J. M. Dekker, A. Fritsche et al., "Variants of CDKAL1 and IGF2BP2 affect first-phase insulin secretion during hyperglycaemic clamps," *Diabetologia*, vol. 51, no. 9, pp. 1659–1663, 2008.

[106] L. Pascoe, A. Tura, S. K. Patel et al., "Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic β-cell function," *Diabetes*, vol. 56, no. 12, pp. 3101–3104, 2007.

[107] C. Dina, D. Meyre, S. Gallina et al., "Variation in FTO contributes to childhood obesity and severe adult obesity," *Nature Genetics*, vol. 39, no. 6, pp. 724–726, 2007.

[108] T. M. Frayling, N. J. Timpson, M. N. Weedon et al., "A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity," *Science*, vol. 316, no. 5826, pp. 889–894, 2007.

[109] E. Zeggini, L. J. Scott, R. Saxena, and B. F. Voight, "Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes," *Nature Genetics*, vol. 40, no. 5, pp. 638–645, 2008.

[110] N. Grarup, G. Andersen, N. T. Krarup et al., "Association testing of novel type 2 diabetes risk alleles in the JAZF1,CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 Loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged danes," *Diabetes*, vol. 57, no. 9, pp. 2534–2540, 2008.

[111] A. M. Simonis-Bik, G. Nijpels, T. W. van Haeften et al., "Gene variants in the novel type 2 diabetes loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B affect different aspects of pancreatic β-cell function," *Diabetes*, vol. 59, no. 1, pp. 293–301, 2010.

[112] A. Jonsson, C. Ladenvall, T. S. Ahluwalia et al., "Effects of common genetic variants associated with type 2 diabetes and glycemic traits on alpha- and beta-cell function and insulin action in humans," *Diabetes*, vol. 62, no. 8, pp. 2978–2983, 2013.

[113] K. Yasuda, K. Miyake, Y. Horikawa et al., "Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus," *Nature Genetics*, vol. 40, no. 9, pp. 1092–1097, 2008.

[114] H. Unoki, A. Takahashi, T. Kawaguchi et al., "SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations," *Nature Genetics*, vol. 40, no. 9, pp. 1098–1102, 2008.

[115] J. Rung, S. Cauchi, A. Albrechtsen et al., "Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia," *Nature Genetics*, vol. 41,no. 10, pp. 1110–1115, 2009.

[116] B. F. Voight, L. J. Scott, V. Steinthorsdottir et al., "Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis," *Nature Genetics*, vol. 42, no. 7, pp. 579–589, 2010.

[117] F.-J. Tsai, C.-F. Yang, C.-C. Chen et al., "A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese," *PLoS Genetics*, vol. 6, no. 2, Article ID e1000847, 2010.

[118] L. Qi, M. C. Cornelis, P. Kraft et al., "Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes," *Human Molecular Genetics*, vol. 19, no. 13, pp. 2706–2715, 2010.

[119] T. Yamauchi, K. Hara, S. Maeda 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*, vol. 42, no. 10, pp. 864–868, 2010.

[120] X. O. Shu, J. Long, Q. Cai et al., "Identification of new genetic risk variants for Type 2 Diabetes," *PLoS Genetics*, vol. 6, no. 9, Article ID e1001127, 2010.

[121] C. W. McDonough, N. D. Palmer, P. J. Hicks et al., "A genome-wide association study for diabetic nephropathy genes in African Americans," *Kidney International*, vol. 79, no. 5, pp. 563–572, 2011.

[122] E. Ingelsson, C. Langenberg, M. F. Hivert et al., "Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans," *Diabetes*, vol. 59, no. 5, pp. 1266–1275, 2010.

[123] T. O. Kilpeläinen, M. C. Zillikens, A. Stančákova et al., "Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile," *Nature Genetics*, vol. 43, no. 8, pp. 753–760, 2011.

[124] X. Sim, R. T.-H. Ong, C. Suo et al., "Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia," *PLoS Genetics*, vol. 7, no. 4, Article ID e1001363, 2011.

[125] H. Wang, P. Maechler, P. A. Antinozzi, K. A. Hagenfeldt, and C. B. Wollheim, "Hepatocyte nuclear factor 4α regulates the expression of pancreatic β-cell genes implicated in glucose metabolism and nutrient-induced insulin secretion," *Journal of Biological Chemistry*, vol. 275, no. 46, pp. 35953–35959, 2000.

[126] I. M. Heid, A. U. Jackson, J. C. Randall et al., "Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution," *Nature Genetics*, vol. 42, no. 11,pp. 949–960, 2010.

[127] Y. S. Cho, C. H. Chen, C. Hu et al., "Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians," *Nature Genetics*, vol. 44, no. 1, pp. 67–72, 2012.

[128] Y. Yang, B. H.-J. Chang, V. Yechoor et al., "The Krüppel-like zinc finger protein GLIS3 transactivates neurogenin 3 for proper fetal pancreatic islet differentiation in mice," *Diabetologia*, vol. 54, no. 10, pp. 2595–2605, 2011.

[129] S. H. Kwak and K. S. Park, "Genetics of type 2 diabetes and potential clinical implications," *Archives of Pharmacal Research*, vol. 36, no. 2, pp. 167–177, 2013.

[130] B. F. Voight, H. M. Kang, J. Ding et al., "Correction: the metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits," *PLoS Genet*, vol. 9, no. 4, 2013.

[131] A. P. Morris, B. F. Voight, T. M. Teslovich et al., "Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes," *Nature Genetics*, vol. 44, no. 9, pp. 981–990, 2012.

[132] R. A. Scott, V. Lagou, R. P. Welch et al., "Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways," *Nature Genetics*, vol. 44, no. 9, pp. 991–1005, 2012.

[133] S. M. Purcell, N. R. Wray, J. L. Stone et al., "Common polygenic variation contributes to risk of schizophrenia and bipolar disorder," *Nature*, vol. 460, no. 7256, pp. 748–752, 2009.

[134] J. Yang, B. Benyamin, B. P. McEvoy et al., "Common SNPs explain a large proportion of the heritability for human height," *Nature Genetics*, vol. 42, no. 7, pp. 565–569, 2010.

[135] E. A. Stahl, D. Wegmann, G. Trynka et al., "Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis," *Nature Genetics*, vol. 44, no. 5, pp. 483–489, 2012.

[136] M. I. McCarthy, "Casting a wider net for diabetes susceptibility genes," *Nature Genetics*, vol. 40, no. 9, pp. 1039–1040, 2008.

[137] M. C. Y. Ng, K. S. Park, B. Oh et al., "Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians," *Diabetes*, vol. 57, no. 8, pp. 2226–2233, 2008.

[138] M. Imamura, S. Maeda, T. Yamauchi et al., "A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations," *Human Molecular Genetics*, vol. 21, no. 13, pp. 3042–3049, 2012.

[139] R. Chen, E. Corona, M. Sikora et al., "Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases," *PLoS Genetics*, vol. 8, no. 4, Article ID e1002621, 2012.

[140] L. Carulli, S. Rondinella, S. Lombardini, I. Canedi, P. Loria, and N. Carulli, "Review article: diabetes, genetics and ethnicity," *Alimentary Pharmacology & Therapeutics*, vol. 22, supplement 2, pp. 16–19, 2005.

[141] J. V. Neel, "Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? 1962," *Bulletin of the World Health Organization*, vol. 77, no. 8, pp. 694–693, 1999.

[142] N. R. Sloan, "Ethnic distribution of diabetes mellitus in Hawaii," *Journal of the American Medical Association*, vol. 183, pp. 419–424, 1963.

[143] J. T. Tan, D. P. K. Ng, S. Nurbaya et al., "Polymorphisms identified through genome-wide association studies and their associations with type 2 diabetes in Chinese, Malays, and Asian-Indians in Singapore," *Journal of Clinical Endocrinology and Metabolism*, vol. 95, no. 1, pp. 390–397, 2010.

[144] F. Takeuchi, M. Serizawa, K. Yamamoto et al., "Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population," *Diabetes*, vol. 58, no. 7, pp. 1690–1699, 2009.

[145] B. Cui, X. Zhu, M. Xu et al., "A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese," *PLoS ONE*, vol. 6, no. 7, Article ID e22353, 2011.

[146] H. Li, W. Gan, L. Lu et al., "A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans," *Diabetes*, vol. 62, no. 1, pp. 291–298, 2013.

[147] A. N. Kho, M. G. Hayes, L. Rasmussen-Torvik et al., "Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study," *Journal of the American Medical Informatics Association*, vol. 19, no. 2, pp. 212–218, 2012.

[148] N. D. Palmer, C. W. McDonough, P. J. Hicks et al., "A genome-wide association search for type 2 diabetes genes in African Americans," *PLoS ONE*, vol. 7, no. 1, Article ID e29202, 2012.

[149] E. J. Parra, J. E. Below, S. Krithika 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*, vol. 54, no. 8, pp. 2038–2046, 2011.

[150] J. E. Below, E. R. Gamazon, J. V. Morrison et al., "Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals," *Diabetologia*, vol. 54, no. 8, pp. 2047–2055, 2011.

[151] R. Tabassum, G. Chauhan, O. P. Dwivedi et al., "Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21," *Diabetes*, vol. 62, no. 3, pp. 977–986, 2013.

[152] R. Saxena, D. Saleheen, L. F. Been et al., "Genome-wide association study identifies a novel locus contributing to type

2 diabetes susceptibility in sikhs of punjabi origin from India," *Diabetes*, vol. 62, no. 5, pp. 1746–1755, 2013.

[153] M. S. Sandhu, M. N. Weedon, K. A. Fawcett et al., "Common variants in WFS1 confer risk of type 2 diabetes," *Nature Genetics*, vol. 39, no. 8, pp. 951–953, 2007.

[154] A. Kong, V. Steinthorsdottir, G. Masson et al., "Parental origin of sequence variants associated with complex diseases," *Nature*, vol. 462, no. 7275, pp. 868–874, 2009.

[155] J. Dupuis, C. Langenberg, I. Prokopenko et al., "New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk," *Nature Genetics*, vol. 42, no. 2, pp. 105–116, 2010.

[156] L. B. Jorde, "Linkage disequilibrium and the search for complex disease genes," *Genome Research*, vol. 10, no. 10, pp. 1435–1444, 2000.

[157] N. A. Rosenberg, L. Huang, E. M. Jewett, Z. A. Szpiech, I. Jankovic, and M. Boehnke, "Genome-wide association studies in diverse populations," *Nature Reviews Genetics*, vol. 11,no. 5, pp. 356–366, 2010.

[158] X. Ke, "Presence of multiple independent effects in risk loci of common complex human diseases," *The American Journal of Human Genetics*, vol. 91, no. 1, pp. 185–192, 2012.

[159] A. Gusev, G. Bhatia, N. Zaitlen et al., "Quantifying missing heritability at known GWAS loci," *PLOS Genetics*, vol. 9, no. 12, Article ID e1003993, 2013.

[160] L. Qi, M. C. Cornelis, C. Zhang, R. M. van Dam, and F. B. Hu, "Genetic predisposition, Western dietary pattern, and the risk of type 2 diabetes in men," *American Journal of Clinical Nutrition*, vol. 89, no. 5, pp. 1453–1458, 2009.

[161] J. C. Florez, K. A. Jablonski, N. Bayley et al., "TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program," *New England Journal of Medicine*, vol. 355, no. 3, pp. 241–250, 2006.

[162] E. Fisher, H. Boeing, A. Fritsche, F. Doering, H.-G. Joost, and M. B. Schulze, "Whole-grain consumption and transcription factor-7-like 2 (TCF7L2) rs7903146: gene-diet interaction in modulating type 2 diabetes risk," *British Journal of Nutrition*, vol. 101,no. 4, pp. 478–481,2009.

[163] L. Qi and J. Liang, "Interactions between genetic factors that predict diabetes and dietary factors that ultimately impact on risk of diabetes," *Current Opinion in Lipidology*, vol. 21, no. 1, pp. 31–37, 2010.

[164] P. Shetty, "Public health: India's diabetes time bomb," *Nature*, vol. 485, no. 7398, pp. S14–S16, 2012.

[165] D. J. P. Barker, C. N. Hales, C. H. D. Fall, C. Osmond, K. Phipps, and P. M. S. Clark, "Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth," *Diabetologia*, vol. 36, no. 1, pp. 62–67, 1993.

[166] P. H. Whincup, S. J. Kaye, C. G. Owen et al., "Birth weight and risk of type 2 diabetes a systematic review," *Journal of the American Medical Association*, vol. 300, no. 24, pp. 2886–2897, 2008.

[167] M. van Hoek, J. G. Langendonk, S. R. de Rooij, E. J. G. Sijbrands, and T. J. Roseboom, "Genetic variant in the IGF2BP2 gene may interact with fetal malnutrition to affect glucose metabolism," *Diabetes*, vol. 58, no. 6, pp. 1440–1444, 2009.

[168] S. R. de Rooij, R. C. Painter, D. I. W. Phillips et al., "The effects of the Pro12Ala polymorphism of the peroxisome proliferator-activated receptor-γ2 gene on glucose/insulin metabolism interact with prenatal exposure to famine," *Diabetes Care*, vol. 29, no. 5, pp. 1052–1057, 2006.

[169] N. Pulizzi, V. Lyssenko, A. Jonsson et al., "Interaction between prenatal growth and high-risk genotypes in the development of type 2 diabetes," *Diabetologia*, vol. 52, no. 5, pp. 825–829, 2009.

[170] V. Steinthorsdottir, G. Thorleifsson, I. Reynisdottir et al., "A variant in CDKAL1 influences insulin response and risk of type 2 diabetes," *Nature Genetics*, vol. 39, no. 6, pp. 770–775, 2007.

[171] J. R. B. Perry, B. F. Voight, L. Yengo 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 Genetics*, vol. 8, no. 5, Article ID e1002741, 2012.

[172] R. C. Ma, C. Hu, C. H. Tam et al., "Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4," *Diabetologia*, vol. 56, no. 6, pp. 1291–1305, 2013.

[173] J. Huang, D. Ellinghaus, A. Franke, B. Howie, and Y. Li, "1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data," *European Journal of Human Genetics*, vol. 20, no. 7, pp. 801–805, 2012.

[174] J. T. Salonen, P. Uimari, J.-M. Aalto et al., "Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium," *American Journal of Human Genetics*, vol. 81, no. 2, pp. 338–345, 2007.

[175] S. Maeda, N. Osawa, T. Hayashi, S. Tsukada, M. Kobayashi, and R. Kikkawa, "Genetic variations associated with diabetic nephropathy and type II diabetes in a Japanese population," *Kidney International*, vol. 72, no. 106, pp. S43–S48, 2007.

[176] L. R. Pasquale, S. J. Loomis, H. Aschard et al., "Exploring genome-wide—dietary heme iron intake interactions and the risk of type 2 diabetes," *Frontiers in Genetics*, vol. 4, article 7, 2013.

[177] S. H. Kwak, S.-H. Kim, Y. M. Cho et al., "A genome-wide association study of gestational diabetes mellitus in Korean women," *Diabetes*, vol. 61, no. 2, pp. 531–541, 2012.

[178] D. Zabaneh and D. J. Balding, "A genome-wide association study of the metabolic syndrome in Indian Asian men," *PLoS ONE*, vol. 5, no. 8, Article ID e11961, 2010.

CHAPTER 3: CHARACTERISING THE GENETIC RISK FOR TYPE 2 DIABETES IN A MALAYSIAN MULTI-ETHNIC COHORT

Abdullah N, Abdul Murad N. A, Attia J, Oldmeadow C, Mohd Haniff E. A, Syafruddin S. E, Abd Jalal N, Ismail N, Ishak M, Jamal R, Scott R.J, Holliday E. G

Diabetes Medicine. 2015 Oct;32 (10):1377-84. doi: 10.1111/dme.12735. Epub 2015 Mar 24 (PMID: 25711284)

3.1 Statement of Co-authors

"As co-authors of the paper:
Abdullah N, Abdul Murad N. A, Attia J, Oldmeadow C, Mohd Haniff E. A, Syafruddin S. E, Abd Jalal N, Ismail N , Ishak M , Jamal R, Scott R. J, Holliday E . G, Characterising the genetic risk for type 2 diabetes in a Malaysian multi-ethnic cohort. Diabetes Medicine. 2015 Oct;32 (10):1377-84. doi: 10.1111/ dme.12735. Epub 2015 Mar 24 (PMID: 25711284), we confirm that Noraidatulakma Abdullah contributed to this publication by performing the analysis, interpreting the result and writing the manuscript."

Co-Authors:

Dr Elizabeth Holliday
Signed
Date: 31 July 2017

Professor John Attia
Signed
Date: 1st August 2017

Dr Christopher Oldmeadow
Signed
Date: 1st August 2017

Professor Rodney Scott
Signed
Date: 2/8/2017

Professor Datuk A. Rahman A. Jamal
Signed
Date:

Dr Nor Azian Abdul Murad
Signed
Date:

Dr Ezanee Azlina Mohd Haniff
Signed
Date:

Saiful Effendi Syafruddin

Signed

Date:

Nazihah Abd Jalal
Signed

Date:

Norliza Ismail

Signed

Date:

Muhiddin Ishak

Signed

Date:

Faculty Assistant Dean (Research Training)

Professor Robert Callister

Signed

Date: 3/08/2017

3.2 Summary of Publication 2

Introduction

The prevalence of T2D in Malaysia is rising rapidly, with Malaysia having one of the highest comparative prevalences of T2D among Asian countries. In spite of this, T2D in the Malaysian population remains relatively understudied. Malaysia is a multi-ethnic country with three major ancestral groups: Malays, Chinese and Indians. The prevalence of T2D varies among these ancestral groups in spite of them sharing the same environment, being highest in Indian, intermediate in Malays and lowest in Chinese. It is unknown to what extent these prevalence differences between these groups, and between the Malaysian and other global populations, may reflect differences in the frequency or effect of genetic risk alleles. Although T2D has a substantial genetic component, many large scales genetic studies of T2D have been conducted in populations of European ancestry, with few, if any, conducted in Malaysia.

This study represents the first detailed genetic study of T2D conducted in Malaysia. It assessed the contribution of 62 T2D genetic risk alleles to T2D in the three Malaysian ancestral groups. Association between known genetic variants and T2D were assessed individually and in combination via a genetic risk score both within and across ancestral groups. We also estimated the T2D risk variation explained by these genetic variants, and compared this between ancestral groups

Characterising the genetic risk for type 2 diabetes in a Malaysian multi-ethnic cohort

Running head: Type 2 diabetes genetics in Malaysia

N. Abdullah [1,2], N. A. Abdul Murad [2], J. Attia [3,4], C. Oldmeadow [3,4], E. A. Mohd Haniff [2], S. E. Syafruddin [2], N. Abd Jalal [2], N. Ismail [2], M. Ishak [2], R. Jamal [2], R.J. Scott [1,5], E. G. Holliday [3,4]

[1] School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia

[2] UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[3] Clinical Research Design, IT and Statistical Support (CReDITSS) Unit, Hunter Medical Research Institute, Newcastle, New South Wales, Australia

[4] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia

[5] Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales, Australia

Corresponding Authors:

Elizabeth G. Holliday

Mailing address:

CReDITSS, HMRI

C/- University Drive

Callaghan NSW 2308

Australia

E-mail: Liz.Holliday@newcastle.edu.au

Telephone: +61(2)40420508

Facsimile: +61(2)40420039

Rahman Jamal

Mailing address:

UKM Medical Molecular Biology Institute (UMBI)

Jalan Yaacob Latif, Bandar Tun Razak, Cheras, 56000 Kuala Lumpur, Malaysia

E-mail: rahmanj@ppukm.ukm.edu.my

Telephone: +60(3)91459000

Facsimile: +60(3)91717185

Word count: 3,207

Funding Sources

NA is supported by Ministry of Higher Education of Malaysia and Universiti Kebangsaan Malaysia. EGH was supported by a Fellowship (100071) from the Australian Heart Foundation and National Stroke Foundation. The Malaysian Cohort Study is funded by a top-down grant from the National Biotechnology Division, Ministry of Science, Technology and Innovation (MOSTI), Malaysia (ER-05-01-02-MEB001).

Conflict of Interest

We declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Novelty statement

- This is the first large-scale genetic study of type 2 diabetes in a Malaysian population
- We assessed 62 SNPs previously associated with type 2 diabetes for their association with disease in 2,392 type 2 diabetes cases and 2,594 controls of Malay, Chinese and Indian ancestry
- Seven individual SNPs were associated with type 2 diabetes after multiple testing adjustment
- We observed highly significant excess in concordance of allelic effect directions between Malaysian and previously studied populations (P=1x10^{-8})
- A genetic risk score including the 62 SNPs showed strong association in the Malaysian sample (P=2x10^{-16}) and explained 1.0-1.7% of disease variance

Abstract

Aims: While genome-wide association studies (GWAS) have identified numerous type 2 diabetes risk variants across diverse populations, the Malaysian population remains unstudied to date. We characterised the association of known type 2 diabetes risk variants in Malaysian subjects of Malay, Chinese and Indian ancestry from The Malaysian Cohort project.

Methods: Using the MetaboChip array, 1,604 Malays (722 cases, 882 controls), 1,654 Chinese (819 cases, 835 controls) and 1,728 Indians (851 cases, 877 controls) were genotyped. First, 62 candidate SNPs previously associated with type 2 diabetes were assessed for association via logistic regression within ancestral groups, and then across ancestral groups via meta-analysis. Second, estimated odds ratios were assessed for excess directional concordance with previously studied populations. Third, a genetic risk score (GRS) aggregating allele dosage across the candidate SNPs was tested for association within and across ancestral groups.

Results: After Bonferroni correction, 7 individual SNPs were associated with type 2 diabetes in the combined Malaysian sample. We observed a highly significant excess in concordance of effect directions between Malaysian and previously studied populations. The GRS was strongly associated with type 2 diabetes in all Malaysian groups, explaining from 1.0 to 1.7% of total type 2 diabetes risk variance.

Conclusion: This study suggests substantial overlap of the genetic risk alleles underlying type 2 diabetes in Malaysian and other populations.

Introduction

Type 2 diabetes is continuing to grow in incidence and prevalence worldwide. In 2013, there were 382 million people worldwide living with diabetes and this number is projected to escalate by 55%, particularly in low and middle income countries (1). In Southeast Asia alone, the incidence of type 2 diabetes is projected to increase by 71% by 2035 (1). According to data from International Diabetes Federation (IDF), Malaysia has the highest comparative prevalence of type 2 diabetes among Asian countries in 2013 (1). Malaysia is a multi-ethnic country whose population of 28.3 million (2) includes three major ancestral groups: Malays (~63%); Chinese (~25%); and Indian (~7%). The prevalence of type 2 diabetes between Malaysian populations appears to differ among the three major groups, with Asian Indians having the highest prevalence (25% to 28%), followed by Malays (17% to 19%) and the lowest apparent prevalence in Chinese (9% to 14%) (3).

Type 2 diabetes has a substantial genetic component and genome-wide association studies (GWAS) have identified more than 100 individual genetic variants associated with the condition. However, the majority of type 2 diabetes GWAS have been conducted in populations of European ancestry, which contribute only a fraction of total human genetic variation. Recent studies in broader populations show the importance of extending the population base of GWAS for type 2 diabetes. Benefits include the potential discovery of novel risk alleles due to population allele frequency differences or population specificity of risk alleles, improved fine-mapping due to population differences in linkage disequilibrium (4) and the ability to characterise transferability and consistency of risk alleles across populations (4). For a range of known risk alleles, initial analyses suggest population overlap of individually associated variants and consistency of allelic effects, both in direction and magnitude (4, 5). The largest trans-ethnic GWAS of type 2 diabetes included 26,488 cases and 83,964 controls from previously published type 2 diabetes GWAS samples of European,

East Asian, South Asian, Mexican and Mexican American ancestry (4). It detected seven new susceptibility loci and a significant excess in directional consistency of risk alleles across populations, indicating relevance of established risk loci across diverse ancestral groups.

To date, there are no published type 2 diabetes GWAS from the Malaysian population, in spite of the high comparative type 2 diabetes prevalence in this country. We conducted the first large-scale genetic study of type 2 diabetes using samples selected from The Malaysian Cohort project (3). Using genotype data generated using the Illumina Metabochip array, we sought to characterise the association of known type 2 diabetes loci in Malaysian samples of Malay, Chinese and Indian ancestry. Our study had three principal aims: i) to assess the association of individual, previously reported type 2 diabetes risk variants with type 2 diabetes within and across Malaysian ancestry groups; ii) to assess evidence for excess concordance in the directional effect of type 2 diabetes risk alleles between previously studied and the Malaysian population, and; iii) to test genetic risk scores that combine information across multiple SNPs for association with type 2 diabetes in Malaysian groups.

Methods

Data sources and study samples

This was a nested case-control study. Type 2 diabetes cases and controls of Malay, Chinese and Indian ancestry were selected from the Malaysian Cohort (TMC), a prospective population-based cohort including 106,527 volunteers aged between 35 and 70 years. Subjects were recruited between April 2006 and September 2012 from regions across Malaysia (3). Comprehensive baseline measurements included fasting plasma glucose (FPG). For the current study, samples with FPG exceeding 7.5 mmol/L (or 126 mg/dL) were classified as type 2 diabetes, while controls had FPG <5.5 mmol/L (or 99 mg/dL).

A total of 1,604 Malays (722 cases, 882 controls), 1,654 Chinese (819 cases, 835 controls) and 1,728 Indians (851 cases, 877 controls) were selected for genotyping. For selection, ethnicity was defined using the self- reported ethnicity of the subject and their family for the three preceding generations. All relevant ethics approvals for The Malaysia Cohort were approved by the institutional review and ethics board of the Universiti Kebangsaan Malaysia, in accordance with the declaration of Helsinki. All subjects gave written, informed consent for participation in the study.

Genotyping and quality control

Samples were genotyped at the UKM Medical Molecular Biology Institute, Kuala Lumpur, Malaysia using the MetaboChip array (Illumina Inc, USA). This custom genotyping array includes 196,725 variants from loci previously implicated in cardiometabolic disease traits, and provides a high-throughput, cost-effective approach to genotyping SNPs previously associated with type 2 diabetes (6). Genotype calling was performed using Illumina GenomeStudio software with default quality score (GenCall) thresholds of ≥0.3 and ≥0.25 for

overall SNPs and individual genotypes, respectively. Manual quality control (QC) of genotype data was performed using PLINK (7). From the full set of genotyped SNPs, we first excluded SNPs with minor allele frequency (MAF) <0.01, missingness >0.05, or with significant deviation from Hardy-Weinberg equilibrium ($P<10^{-5}$) in any of the three ancestral groups. In addition, cluster plots were visually checked for all selected candidate SNPs (see below) to ensure clear separation of genotype clusters (8). We then excluded samples with missingness >0.05, outlying heterozygosity (+/- 8 SD from the mean), discrepant clinical and genotypic gender, accidental duplication, or cryptic relatedness (IBS sharing proportion >0.1875; midway between second and third-degree relatives). Genetic ancestry was assessed by principal components analysis (PCA) using reference data from the Singaporean Genome Variation Project (SGVP) and EIGENSTRAT software (9). The SGVP was used due to high similarity between the Singaporean and Malaysian populations. The SGVP includes reference genotype data for 89 Singaporean Malays, 96 Singaporean Chinese and 83 Singaporean Indians. Malaysian Cohort samples not clustering with their specified ancestral group (± 6 SD from the cluster mean on the first two principal components) were removed.

After performing SNP- and sample-level quality control, we performed logistic regression of case-control status against allelic dose across all remaining SNPs within each ancestral group. These analyses were performed with sequential adjustment for up to ten principal components (PCs) to calculate genomic inflation factors (λ_{GC}) and inform decisions about PC inclusion in candidate SNP association models, in order to minimise λ_{GC}.

Candidate SNP Selection

More than 40 GWAS studies of type 2 diabetes and its complications have been published and listed in the online Catalogue of published Genome-Wide Association Studies (10). Using the Catalogue and a recent comprehensive review of type 2 diabetes genetic

associations (11), we identified SNPs previously showing genome-wide significant association ($P<5 \times 10^{-8}$) with type 2 diabetes. Identified SNPs were selected for testing in our Malaysian sample if they were: a) present on the Metabochip array, and; b) passed quality control in at least two of the three Malaysian population groups. For type 2 diabetes-associated loci containing multiple associated SNPs, we selected a single lead SNP. The final set of candidate SNPs were in approximate linkage equilibrium, with all pairwise r^2 <0.5 based on linkage disequilibrium in HapMap Chinese/Japanese combined reference data (CHB/JPT) (12).

Statistical Analyses

Association of each candidate SNP with type 2 diabetes was first assessed separately within ancestral groups using a logistic model assuming an additive allelic effect on the log-odds scale. Principal components of ancestry were included as covariates as indicated for each ancestral group. If the genomic inflation factor (λ_{GC}) exceeded 1, the standard errors of estimated SNP allele coefficients were further adjusted via genomic control, based on the observed λ_{GC} for the relevant ancestral group (13). Association summary statistics from each ancestral group were combined via inverse-variance weighted, fixed-effects meta-analysis using METAL (14). Heterogeneity of allelic effects was assessed using Cochran's Q Statistic. The experiment-wide significance level was derived via Bonferroni correction for the number of lead SNPs assessed via meta-analysis. Quantile-quantile (Q-Q) plots were also generated to visually assess enrichment for true associations.

To supplement individual association tests, the full set of candidate SNPs was assessed for the extent of concordance of allelic effect direction with previous studies, as previously described (4). This was performed both within and across the three ancestral groups. The observed proportion of directionally concordant SNPs was compared with that expected by

chance, with the null proportion being 0.5 for tests within individual ancestral groups, and 0.125 (0.5×0.5×0.5) for the meta-analysis analysis across groups, for which concordance across all groups was required. Observed and expected proportions were compared using a binomial test.

To assess the association evidence aggregated across candidate SNPs, a genetic risk score (GRS) was constructed. The GRS was formed as the weighted sum of reference alleles for each candidate SNP, with weights specified as the log odds ratio (beta coefficient) reported in the original publication. If multiple studies had reported genome-wide significant association of a SNP, we used the effect estimate from the largest study. Scoring was performed using PLINK. Association of the GRS with type 2 diabetes was assessed within each ancestral group via logistic regression. The proportion of case-control variance explained by the score was estimated using Nagelkerke's pseudo R^2. Association evidence for the GRS was combined across ancestral groups via fixed-effects meta-analysis. Association testing was performed using Stata (15).

Secondary analysis

As a secondary analysis, we assessed association with T2D for all Metabochip SNPs passing quality control. Logistic regression within ethnic groups and meta-analysis of results across the three ethnic groups were performed as described for candidate SNPs.

Results

After quality control, 4077 samples remained: 1,323 Malays samples (600 cases, 723 controls), 1344 Chinese samples (654 cases, 690 controls) and 1,410 Indians sample (708 cases, 702 controls). The Malay, Chinese and Indian groups were clearly separated on the first two ancestral principal components (**Supplementary Figure 1**) and each clustered closely with its respective Singaporean group. Based on PCA results and observed genomic inflation factors, the first 3 principal components were selected for inclusion as model covariates in both Malay and Indian groups, to minimise the λ_{GC}. Using logistic models including PCs as covariates, the observed inflation factors for Malay and Indian groups were 1.069 and 1.029, respectively. No principal components were necessary in Chinese, for which the unadjusted λ_{GC} was <1 (0.977).

Candidate SNP association tests

Of the identified 188 SNPs previously showing genome-wide association with type 2 diabetes, 97 had data available in at least two of the three ancestral groups. This set included several clusters of SNPs within a single locus. After selecting a single lead SNP for each locus, 62 SNPs remained. Based on Bonferroni correction for 62 SNPs, the pre-specified, adjusted significance threshold was $\alpha = 0.05/62 = 8.06 \times 10^{-4}$. Power to detect associated SNPs was calculated (16), assuming an additive model, perfect linkage disequilibrium (LD) between risk and marker alleles and an adjusted significance threshold of α=0.000806. For a genetic risk ratio of 1.2, we had 38%, 72% and 85-89% power to identify risk alleles with frequency 0.1, 0.2 and 0.3-0.5 respectively. For a true risk ratio of 1.1, power was low, ranging from 4% to 19% across allele frequencies.

Association results for all SNPs, both within ancestral groups and in the meta-analysis across groups, are shown in **Supplementary Table 1**. Of the 62 SNPs, 7 reached $P<8.06 \times 10^{-4}$

(**Table 1**) and 10 reached a nominal significance threshold of $P<0.05$ (**Supplementary Table 1**) in the meta-analysis across groups. The SNPs reaching $P<8.06 \times 10^{-4}$ were rs10965250 within *CDKN2A* ($P=3\times10^{-5}$), rs4607517 within *GCK* ($P=6\times10^{-5}$), rs7903146 within *TCF7L2* ($P=2\times10^{-4}$), rs9939609 within *FTO* ($P=2\times10^{-4}$), rs12970134 within *MC4R* ($P=3\times10^{-4}$), rs11708067 within *ADCY5* ($P=4\times10^{-4}$), and rs1801282 within *PPARG* ($P=7\times10^{-4}$). Variants reaching nominal significance were rs1801214 in *WFS1* ($P=5\times10^{-3}$), rs6931514 in *CDKAL1* ($P=2\times10^{-3}$), rs3802177 in *SLC30A8* ($P= 7\times10^{-3}$), rs2796441 in *TLE1-FAM75D5* ($P=0.03$), rs1111875 in *HHEX - EXOC6* ($P= 1\times10^{-3}$), rs6583826 in *IDE - RPL11P4* ($P=0.02$), rs174550 in *FADS1* ($P=1\times10^{-3}$), rs1552224 in *ARAP1* ($P=0.01$), rs7177055 in *HMG20A-LINGO1* ($P=0.02$) and rs8042680 in *PRC1; LOC100507118* ($P=0.04$). Within individual groups, 7 of the 62 SNPs reached a nominal significance threshold of $P<0.05$ in Malays, 8 of 58 reached $P<0.05$ in Chinese and 9 of 62 SNPs reached $P<0.05$ in Indians (**Supplementary Table 1**). Quantile-quantile (Q-Q) plots representing the *P*-value distribution for association of the 62 SNPs showed considerable deviation from the distribution expected under the null hypothesis, both within groups and in the meta-analysis (**Supplementary Figure 2**). This suggests considerable enrichment for true associations, in spite of relatively few SNPs reaching the adjusted significance threshold.

Concordance in allelic effect directions between Malay and other populations

Within individual groups, we observed evidence of a significant excess in concordance of allelic effect directions with previously reported values. Of 62 SNPs with data available in Malays, 45 (72.6 %) showed effect directions consistent with previous studies, compared to the 50% expected by chance (binomial $P = 3.37 \times 10^{-5}$). Similarly, in Chinese, 45 of 58 SNPs (77.6%) showed concordant effects (binomial $P = 2.31 \times 10^{-7}$) and in Indians 47 of 62 (75.8%) were directionally consistent with previous findings (binomial $P = 1.04 \times 10^{-6}$). In results from the meta-analysis, 83.9% of SNPs showed a consistent summary effect direction

with that previously reported (52 of 62 SNPs; binomial P=1.90 x 10^{-13}) (see also **Figure 1**). Of these 52 SNPs showing the previously reported effect direction, 33 (63.5%) showed an attenuated magnitude of effect compared to the original publication, significantly more than the 50% expected by chance (binomial P=0.02).

From the meta-analyses of results for individual SNPs, 56 SNPs had data available for all three ancestral groups. Of these, 28 (50.0%) showed consistent effect directions across all ancestral groups, significantly more than the proportion expected by chance (12.5%; binomial P=9.97 x 10^{-9}). These results are consistent with the QQ-plots of SNP P-values, indicating enrichment for true associations among the selected candidate SNPs. They also show that effect directions for type 2 diabetes risk alleles in Malay groups are both relatively homogenous between groups and consistent with results in other ancestral populations.

Association of genetic risk scores

The genetic risk score (GRS) included data from 62 candidate SNPs in the Malay and Indian groups and 58 SNPs in Chinese. The GRS showed significant and consistent association with type 2 diabetes in all ancestral groups (Malay: $P = 4.91 \times 10^{-8}$; Chinese: $P = 1.35 \times 10^{-8}$ and Indian: $P = 4.71 \times 10^{-6}$), reaching a higher level of significance in the meta-analysis across groups ($P = 2.2 \times 10^{-16}$) (**Table 2**). The estimated proportion of type 2 diabetes risk variance explained by the GRS was 1.7% in Chinese, 1.6% in Malays, and 1.0% in Indians. There was no evidence of heterogeneity of the GRS effect across ancestral groups (Cochran's I^2=0.0%; P=0.39) (**Figure 2**). The effect direction of the GRS was consistent with prior evidence both within and across ancestral groups, with risk scores reflecting a higher burden of previously reported risk alleles also associating with increased type 2 diabetes risks in Malaysian groups.

Secondary analysis

Supplementary Figure 3 shows the Manhattan plot of Metabochip-wide meta-analysis results for 106, 701 SNPs passing quality control in at least two of the three ancestral groups. The genomic inflation factor was 1.054. We calculated power for a genome-wide significance threshold of $\alpha = 5 \times 10^{-8}$. For a true risk ratio of 1.2, power ranged from 1% to 19% across allele frequencies. For a risk ratio of 1.1, power was 0 across all allele frequencies. No SNP reached $P < 5 \times 10^{-8}$ in the Metabochip-wide analysis. The strongest associations were observed in the *FTO* gene on chromosome 16 ($P = 3.4 \times 10^{-6}$), and on chromosomes 7, 12 and 13 (including variants within the *OGDH* and *DDX56* genes). A total of 5 markers reached $P < 1 \times 10^{-5}$ (**Supplementary Table 2**).

Discussion

To our knowledge, this represents the first detailed genetic study of type 2 diabetes in Malaysia. We assessed association of previously reported, type 2 diabetes-associated variants at more than 60 loci in the three largest Malaysian ancestral groups: Malays, Chinese and Indians. Meta-analyses across groups identified SNPs in seven loci reaching significance after multiple-testing adjustment, at the *TCF7L2*, *CDKN2A*, *FTO*, *PPARG*, *GCK*, *MC4R* and *ADCY5* loci. In addition, 10 additional SNPs reached nominal significance in *WFS1*, *CDKAL1*, *SLC30A8*, *TLE1-FAM75D5*, *HHEX - EXOC6*, *IDE - RPL11P4*, *FADS1*, *ARAP1*, *HMG20A-LINGO1* and *PRC1; LOC100507118*. The majority of these genes are involved in biological pathways influencing diabetes pathophysiology, including pancreatic beta-cell development/function, insulin availability, glucose utilisation, fatty acid concentrations and obesity. While these variants were each initially identified in European ancestry populations, they have each also shown association in broader global populations, including groups of South-Asian and/or East Asian ancestry (17-23). This study confirms their additional involvement in type 2 diabetes in Malaysia.

In this Malaysian sample, we were unable to confirm individual association for many genetic variants previously associated with type 2 diabetes. A likely explanation was insufficient statistical power to identify variants with small individual effect. Power was reduced first by our modest sample size relative to earlier studies by large, international consortia. Second, for the majority of tested variants, estimated odds ratios in our Malaysian sample were small, generally ranging from 1.0 to 1.2. Indeed, we observed a significant excess of variants with smaller effect size in the Malaysian sample compared with the original study. A tendency for lower effect sizes has also been reported in studies of similar populations from Singapore (17, 24). This may be due to the phenomenon known as “winner’s curse”, or upward bias of effect estimates in the initial reporting study. Alternatively, smaller effect sizes may reflect lower

linkage disequilibrium between assessed and underlying causal variants in these South-East Asian populations. Regardless of the cause, attenuated odds ratios will diminish power to detect trait-variant association. Notably, for the seven variants showing significant association with type 2 diabetes, estimated odds ratios were relatively large, ranging from 1.2 to 1.4.

Notwithstanding limited power for testing individual variants, QQ-plots revealed an excess of nominally associated variants compared to chance expectation. Formal tests also showed a significantly elevated number of SNPs whose estimated effect direction was consistent with earlier studies. This suggests that many of the assessed SNPs may well influence type 2 diabetes risks in the Malaysian population and could demonstrate more significant association in larger samples.

The composite genetic risk score also demonstrated highly significant association with type 2 diabetes, both within individual groups and in the meta-analysis across groups, with all scores having an effect direction consistent with earlier studies. This further supports the relevance of many previously-reported type 2 diabetes risk variants in the Malaysian population. Despite this apparent transferability of type 2 diabetes risk alleles into Malaysia, the genetic risk score explained less than 2% of overall type 2 diabetes risk in any individual group. We do acknowledge that our study assessed association for SNPs representing only 97 of an identified 188 variants previously associated with type 2 diabetes. The effect of including SNPs representing the additional 91 variants is unknown, but would likely produce higher estimates of explained variance. If the additional 91 variants explain a similar, additional proportion of risk, the rapidly escalating type 2 diabetes prevalence in Malaysia (1) seems unlikely to result solely from common genetic variants. Recent environmental changes in dietary patterns and physical activity may contribute more substantially. In addition, lifestyle-related factors such BMI, waist-hip circumference or dietary intake of fats/sugar may interact

with genetic risk alleles to further elevate type 2 diabetes risk in the Malaysian population. Future studies of low frequency variants or epigenetic modifications may also reveal genetic factors influencing the rising prevalence of type 2 diabetes in south-east Asian populations.

Funding Sources

NA is supported by Ministry of Higher Education of Malaysia and Universiti Kebangsaan Malaysia. EGH is supported by a Fellowship (100071) from the Australian Heart Foundation and National Stroke Foundation. The Malaysian Cohort Study is funded by a top-down grant from the National Biotechnology Division, Ministry of Science, Technology and Innovation (MOSTI), Malaysia (ER-05-01-02-MEB001).

Conflict of Interest

We declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Acknowledgments

We thank all UKM Medical Molecular Biology Institute (UMBI) and The Malaysian Cohort staff members and research assistants. The voluntary participation of all the subjects is greatly appreciated.

References

1. International Diabetes Federation. IDF Diabetes Atlas 6th Edition2013; 6:[169 p.].

2. Population Distribution and Basic Demographic Characteristics 2010 [Internet]. Department of Statistics Malaysia,. 2011 [cited 12 May 2014].

3. Jamal R. SZ, S.Z., Kamaruddin M.A., Jalal A.N, Ismail N., Kamil N.M, Abdullah N., Baharudin N., Hussin N.H, Othman H., Mahadi N. M., The Malaysian Cohort Group,. Cohort profile: The Malaysian Cohort (TMC) project: a prospective study of non-communicable diseases in a multi-ethnic population. The International Journal of Epidemiology. 2014:9.

4. Replication DIG, Meta-analysis C. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nature genetics. 2014;46(3):234-44.

5. Saxena R, Elbers CC, Guo Y, Peter I, Gaunt TR, Mega JL, et al. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. American journal of human genetics. 2012;90(3):410-25.

6. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. Correction: The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. PLoS genetics. 2013;9(4).

7. 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.

8. Tindall EA, Petersen DC, Nikolaysen S, Miller W, Schuster SC, Hayes VM. Interpretation of custom designed Illumina genotype cluster plots for targeted association studies and next-generation sequence validation. BMC research notes. 2010;3:39.

9. Teo YY, Sim X, Ong RT, Tan AK, Chen J, Tantoso E, et al. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. Genome research. 2009;19(11):2154-62.

10. A Catalog of Published Genome-Wide Association Studies [Internet]. National HUman Genome Research Institute. 2014 [cited April 2014]. Available from: www.genome.gov/gwastudies.

11. Abdullah N. AJ, Oldmeadow C., Scott R.J. & Holliday E.G. The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings. International Journal of Endocrinology. 2014;2014:21.

12. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24(24):2938-9.

13. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55(4):997-1004.

14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.

15. Sterne JAC BM, Egger M. Meta-analysis in Stata. 2001. In: Systematic Reviews in Health Care [Internet]. London: BMJ Publishing Group; [347-69].

16. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nature genetics. 2006;38(2):209-13.

17. Chen Z, Pereira MA, Seielstad M, Koh WP, Tai ES, Teo YY, et al. Joint effects of known type 2 diabetes susceptibility Loci in genome-wide association study of singapore chinese: the singapore chinese health study. PloS one. 2014;9(2):e87762.

18. Rasmussen-Torvik LJ, Guo X, Bowden DW, Bertoni AG, Sale MM, Yao J, et al. Fasting glucose GWAS candidate region analysis across ethnic groups in the Multiethnic Study of Atherosclerosis (MESA). Genetic epidemiology. 2012;36(4):384-91.

19. Hara K, Fujita H, Johnson TA, Yamauchi T, Yasuda K, Horikoshi M, et al. Genome-wide association study identifies three novel loci for type 2 diabetes. Human molecular genetics. 2014;23(1):239-46.

20. Li H, Kilpelainen TO, Liu C, Zhu J, Liu Y, Hu C, et al. Association of genetic variation in FTO with risk of obesity and type 2 diabetes with data from 96,551 East and South Asians. Diabetologia. 2012;55(4):981-95.

21. Xi B, Takeuchi F, Chandak GR, Kato N, Pan HW, Consortium A-TD, et al. Common polymorphism near the MC4R gene is associated with type 2 diabetes: data from a meta-analysis of 123,373 individuals. Diabetologia. 2012;55(10):2660-6.

22. Chan KH, Niu T, Ma Y, You NC, Song Y, Sobel EM, et al. Common genetic variants in peroxisome proliferator-activated receptor-gamma (PPARG) and type 2 diabetes risk among Women's Health Initiative postmenopausal women. The Journal of clinical endocrinology and metabolism. 2013;98(3):E600-4.

23. Rees SD, Hydrie MZ, O'Hare JP, Kumar S, Shera AS, Basit A, et al. Effects of 16 genetic variants on fasting glucose and type 2 diabetes in South Asians: ADCY5 and GLIS3 variants may predispose to type 2 diabetes. PloS one. 2011;6(9):e24710.

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(4):e1001363.

25. 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(7):579-89.

26. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics. 2010;42(2):105-16.

27. 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 genetics. 2012;8(5):e1002741.

28. 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(9):981-90.

Table 1: Association results for candidate SNPs showing significant association with type 2 diabetes in the meta-analysis across Malaysian groups

Lead SNP [1]	Mapped Gene(s)	Region	RA[2]	OA[3]	Malays		Chinese		Indians		Meta-analysis		
					OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P* [4]	EA[5]
rs10965250(25)	*CDKN2B*	9p21.3	G	A	0.82 (0.66-0.98)	0.02[a]	0.79 (0.68-0.92)	$3x10^{-3}$[a]	0.81 (0.61-1.01)	0.04[a]	0.81 (0.73-0.89)	$3x10^{-5}$	A[b]
rs4607517(26)	*GCK - YKT6*	7p13	A	G	1.25 (1.01-1.49)	0.06	1.58 (1.29-1.94)	$1x10^{-5}$[a]	1.01 (0.78-1.25)	0.91	1.30 (1.15-1.48)	$6x10^{-5}$	A[b]
rs7903146(25)	*TCF7L2*	10q25.2	T	C	1.21 (0.93-1.49)	0.19	1.08 (0.70-1.68)	0.72	1.37 (1.21-1.53)	$2x10^{-4}$[a]	1.30 (1.13-1.49)	$2x10^{-4}$	A[b]
rs9939609(27)	*FTO*	16q12.2	A	T	1.32 (1.15-1.49)	$1x10^{-3}$[a]	1.16 (0.93-1.45)	0.18	1.16 (1.01-1.32)	0.06	1.22 (1.10-1.35)	$2x10^{-4}$	A[b]
rs12970134(28)	*MC4R*	18q21.32	A	G	1.19 (0.97-1.41)	0.12	1.17 (0.96-1.44)	0.13	1.28 (1.12-1.44)	$3x10^{-3}$[a]	1.22 (1.10-1.36)	$3x10^{-4}$	A[b]
rs11708067(26)	*ADCY5*	3q21.1	A	G	0.69 (0.32-1.06)	0.05	N/A[c]	N/A[c]	0.74 (0.54-0.93)	$2x10^{-3}$[a]	1.38 (1.15-1.64)	$4x10^{-4}$	A[b]
rs1801282(4)	*PPARG*	3p25.2	A	G	0.77 (0.36-1.18)	0.2	0.93 (0.57-1.52)	0.78	0.67 (0.41-0.92)	$2x10^{-3}$[a]	1.28 (1.03-1.58)	$7x10^{-4}$	A[b]

[1] SNP previously associated with type 2 diabetes or fasting plasma glucose at genome-wide significance ($P<5x10^{-8}$), with original reference. [2] Risk allele from previous study. [3] Other allele from previous study. [4] $P<8.1x10^{-4}$, incorporating adjustment for testing 62 independent SNPs. [5] Effect allele from meta-analysis.
a Denotes SNPs reaching *P*-value <0.05 in individual ancestral groups
[b] Denotes same effect direction as previously reported
[c] Data not available due to MAF <0.01

Table 2: Association between the genetic risk score and type 2 diabetes within ancestral groups and in the meta-analysis across groups

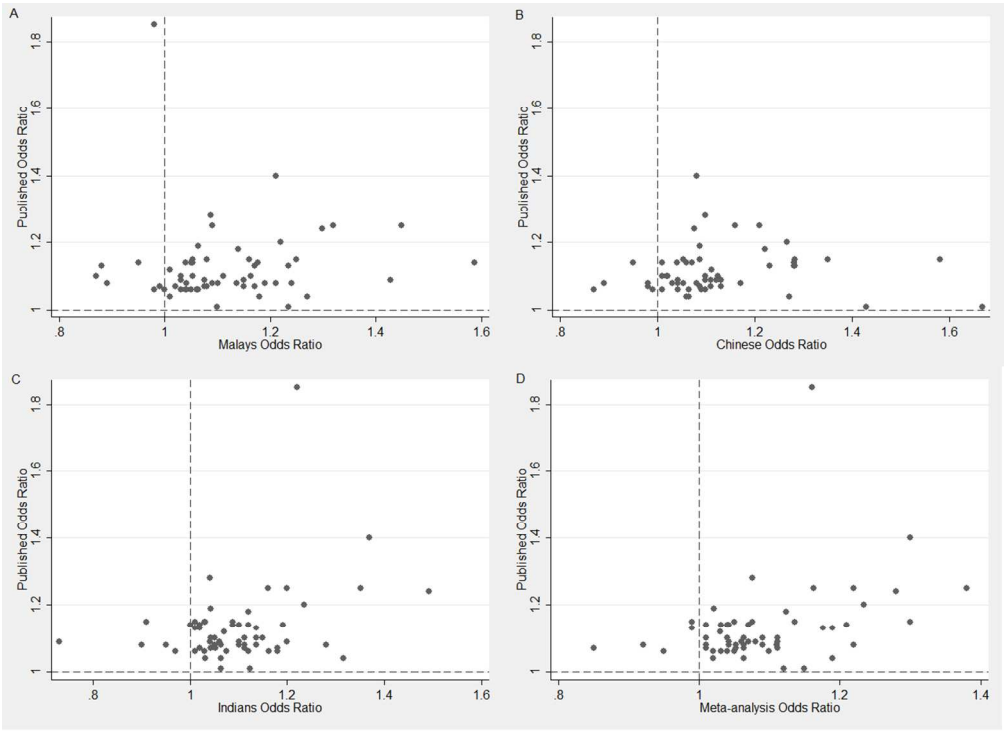
Study	**N_{SNPs}**	***P*-value**	**Effect direction**	**Pseudo R^2**
Malays	62	$4.91x10^{-8}$	+	1.6%
Chinese	58[a]	$1.35x10^{-8}$	+	1.7%
Indians	62	$4.71x10^{-6}$	+	1.0%
Meta-analysis	62	$2.2x10^{-16}$	+	

[a] Number reduced owing to 4 SNPs with MAF<0.01 in Chinese

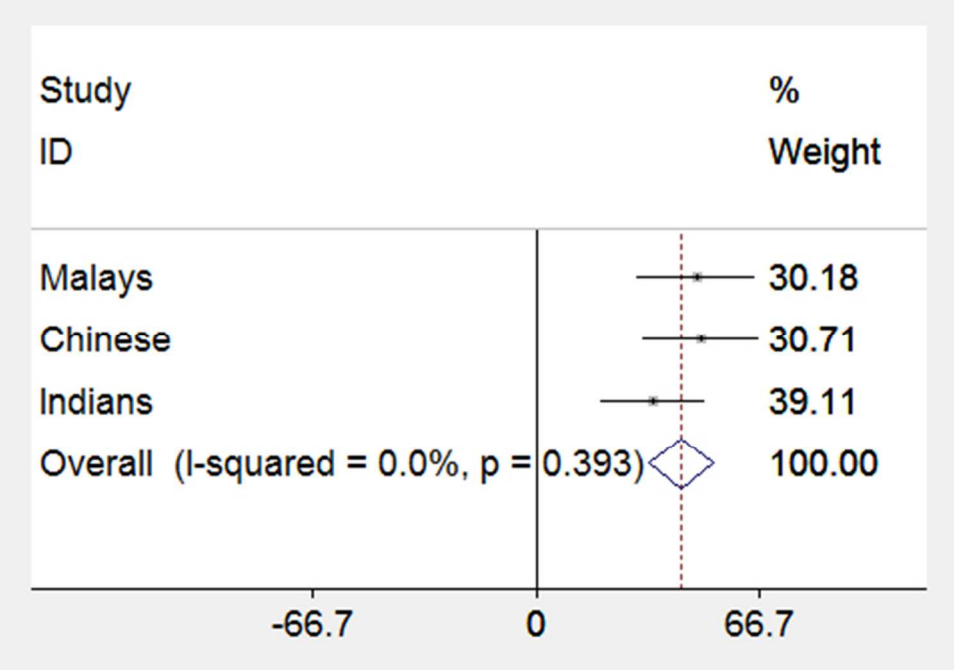
Figure Legends

Figure 1: Bivariate plots comparing odds ratios observed in each Malaysian ancestral group and the meta-analysis across groups, with those previously published. (A) Malays, (B) Chinese, (C) Indians, (D) Combined meta-analysis.

Figure 2: Forest plot showing association of the genetic risk score in the meta-analysis across Malaysian groups.

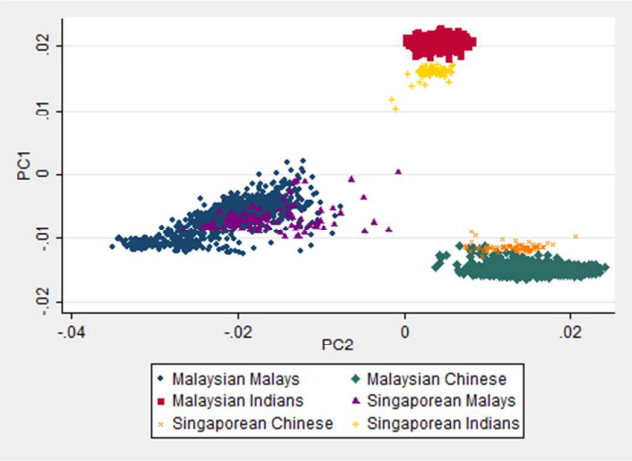


Bivariate plots comparing odds ratios observed in each Malaysian ancestral group and the meta-analysis across groups, with those previously published. (A) Malays, (B) Chinese, (C) Indians, (D) Combined meta-analysis.
180x130mm (180 x 180 DPI)

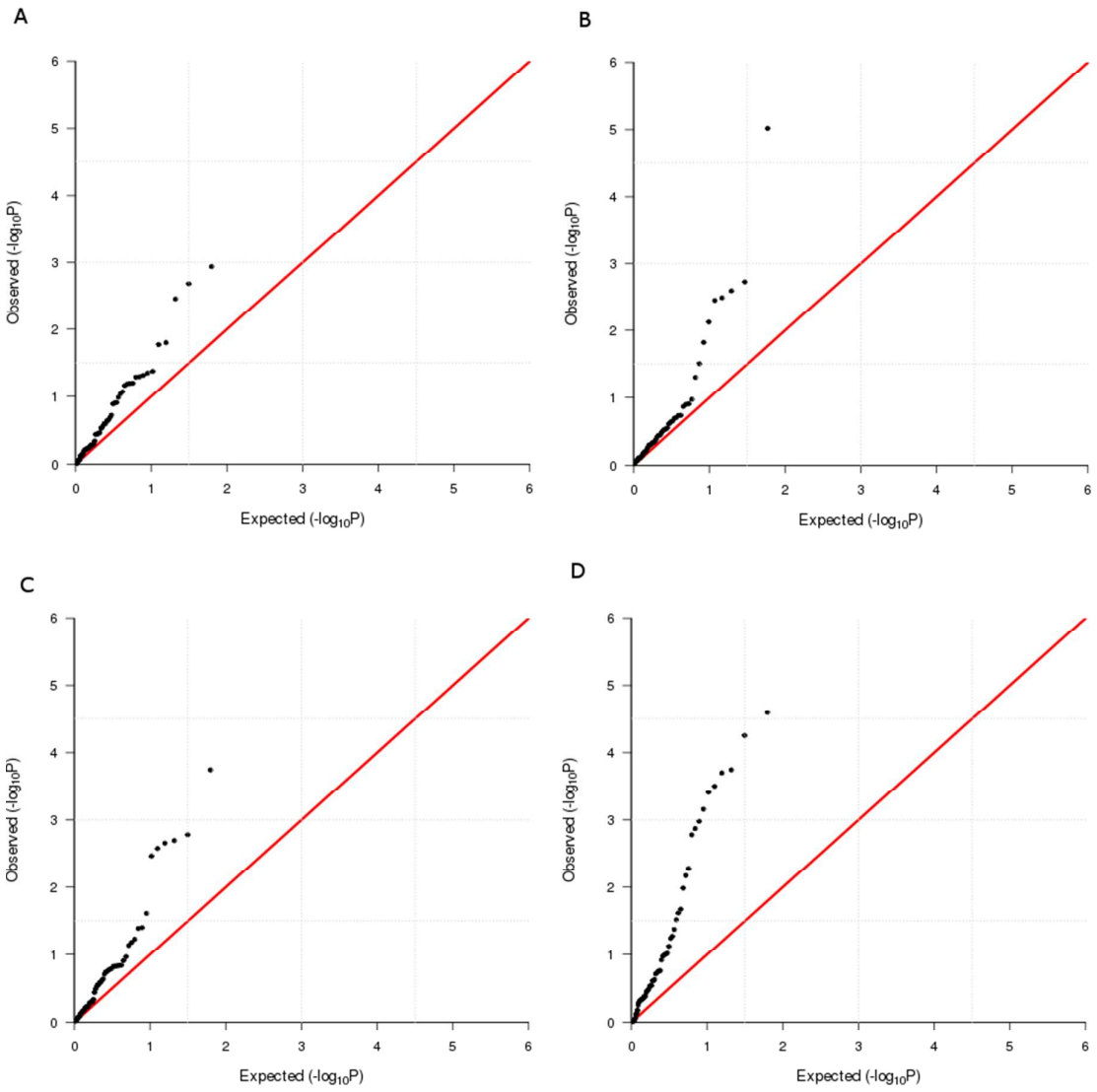


Forest plot showing association of the genetic risk score in the meta-analysis across Malaysian groups.
88x62mm (136 x 136 DPI)

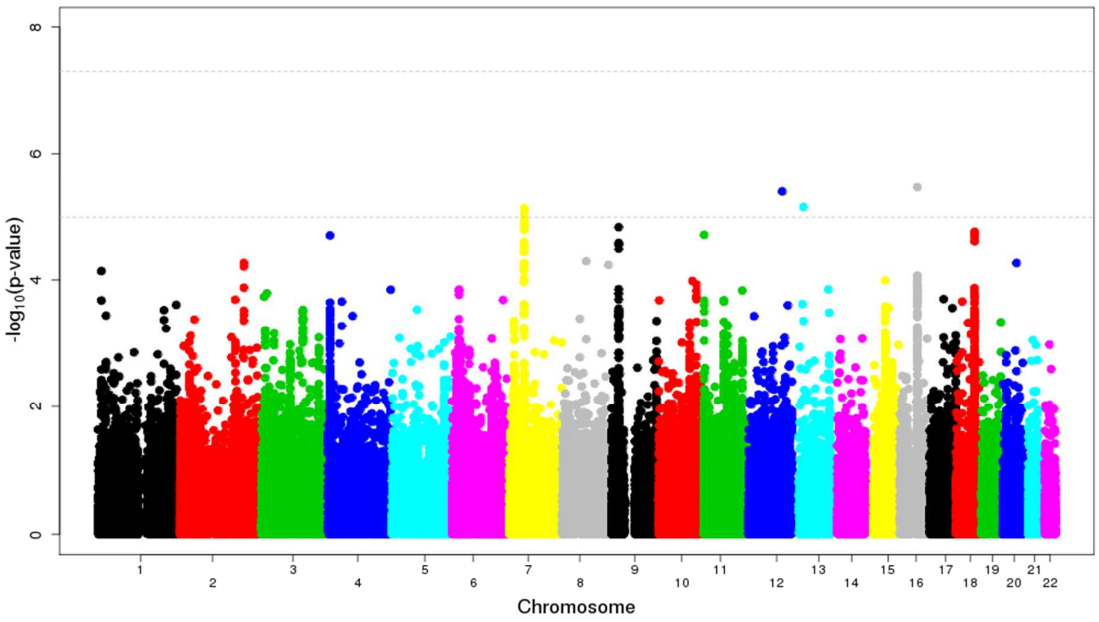
Data Supplement



Supplementary Figure 1: Plot of the first two ancestry principal components in Malaysian samples, incorporating reference data from the Singaporean Genome Variation Project.



Supplementary Figure 2: Quantile-quantile (Q-Q) plots of the *P*-values for association of 62 SNPs observed in each of ethnic groups and meta-analysis. (A) Malays, (B) Chinese, (C) Indians, (D) Combined meta-analysis



Supplementary Figure 3: Manhattan plot of Metabochip-wide meta-analysis results for all SNPs passing quality control in at least two of the three ethnic groups. The genomic inflation factor (λ_{GC}) was 1.054.

Supplementary Table 1: Association summary statistics within and across ancestral groups for 62 autosomal SNPs previously associated with type 2 diabetes

Lead SNPs[1]	Chr:BP	RA[2]/OA[3]	POR[4]	Malays			Chinese			Indians			Meta-analysis		
				OR (95% CI)	*P*	EA[5]	OR (95% CI)	*P*	EA[5]	OR (95% CI)	*P*	EA[5]	OR (95% CI)	*P*	Het *P*	EA[5]
rs17106184(1)	1:50,682,573	G/A	1.1	0.97 (0.68-1.26)	0.83	A[c]	0.89 (0.68-1.17)	0.42	A[c]	0.96 (0.62-1.31)	0.82	A[c]	0.94 (0.79-1.12)	0.46	0.92	A[c]
rs10923931(2)	1:120,319,482	T/G	1.13	0.88 (0.60-1.16)	0.36	A	1.23 (0.79-1.92)	0.35	A[c]	1.02 (0.84-1.19)	0.86	A[c]	0.99 (0.86-1.16)	0.99	0.44	A
rs340874(3)	1:212,225,879	C/T	1.07	0.99 (0.85-1.15)	0.99	G	0.98 (0.84-1.15)	0.84	G	0.98 (0.84-1.13)	0.82	A[c]	1.01 (0.92-1.09)	0.99	0.94	A
rs780094(3)	2:27,594,741	C/T	1.06	0.94 (0.78-1.10)	0.45	A[c]	0.91 (0.79-1.06)	0.22	A[c]	0.86 (0.67-1.05)	0.11	A[c]	0.91 (0.82-1.00)	0.05	0.76	A[c]
rs7578597(4)	2:43,586,327	T/C	1.15	1.08 (0.63-1.53)	0.75	G	1.35 (0.61-3.00)	0.46	G	0.97 (0.75-1.19)	0.79	G[c]	0.99 (0.81-1.21)	0.93	0.71	A
rs13389219(5)	2:165,210,095	A/C	1.09	1.15 (0.91-1.38)	0.25	A[c]	1.13 (0.90-1.41)	0.3	C	0.73 (0.54-0.93)	$2x10^{-3b}$	A	1.07 (0.94-1.21)	0.19	$1x10^{-2}$	A[c]
rs7578326(6)	2:226,801,989	C/T	1.19	0.94 (0.69-1.18)	0.6	A[c]	0.92 (0.68-1.24)	0.57	A[c]	0.96 (0.76-1.17)	0.73	A[c]	0.98 (0.85-1.12)	0.76	0.84	A[c]
rs1801282(1)[a]	3:12,264,800	A/G	1.24	0.77 (0.36-1.18)	0.2	G[c]	0.93 (0.57-1.52)	0.78	G[c]	0.67 (0.41-0.92)	$2x10^{-3b}$	G[c]	1.28 (1.03-1.58)	$7x10^{-4}$	0.8	A[c]
rs4607103(7)	3:64,686,944	C/T	1.09	1.03 (0.86-1.19)	0.74	A	0.96 (0.82-1.13)	0.64	A[c]	1.10 (0.95-1.25)	0.19	G[c]	0.96 (0.87-1.05)	0.36	0.56	A[c]
rs11708067(3)[a]	3:124,548,468	A/G	1.25	0.69 (0.32-1.06)	0.05	G[c]	N/A			0.74 (0.54-0.93)	$2x10^{-3b}$	G[c]	1.38 (1.15-1.64)	$4x10^{-4}$	0.77	A[c]
rs11920090(3)	3:172,200,215	T/A	1.01	0.81 (0.45-1.17)	0.25	A[c]	0.70 (0.25-1.98)	0.5	A[c]	0.89 (0.69-1.08)	0.24	A[c]	0.87 (0.73-1.03)	0.1	0.84	A[c]
rs4402960(8)	3:186,994,381	T/G	1.14	0.95 (0.79-1.12)	0.57	A	1.28 (1.07-1.54)	$8x10^{-3b}$	A[c]	1.02 (0.86-1.17)	0.84	A[c]	1.07 (0.97-1.17)	0.19	0.04	A[c]
rs6808574(1)	3:189,223,217	C/T	1.07	0.87 (0.48-1.27)	0.5	A[c]	N/A			0.96 (0.76-1.15)	0.67	A[c]	0.94 (0.79-1.12)	0.51	0.68	A[c]
rs1801214(6)	4:6,353,923	T/C	1.13	0.81 (0.61-1.02)	0.05	G[c]	0.78 (0.58-1.05)	0.11	G[c]	0.88 (0.70-1.05)	0.15	G[c]	1.19 (1.05-1.35)	$5x10^{-3}$ [b]	0.79	A[c]

rs6813195(1)	4:153,739,925	C/T	1.08	0.96 (0.81-1.11)	0.57	A[c]	0.97 (0.83-1.13)	0.7	A[c]	0.94 (0.79-1.10)	0.46	A[c]	0.96 (0.87-1.05)	0.34	0.97	A[c]
rs702634(1)	5:53,307,177	A/G	1.06	1.00 (0.79-1.21)	0.99	G	0.91 (0.73-1.12)	0.36	G[c]	0.97 (0.80-1.13)	0.67	A	1.04 (0.93-1.17)	0.45	0.81	A[c]
rs459193(5)	5:55,842,508	G/A	1.08	1.10 (0.95-1.26)	0.22	G[c]	0.96 (0.83-1.12)	0.63	A[c]	0.95 (0.79-1.10)	0.51	G	0.94 (0.86-1.03)	0.18	0.87	A[c]
rs4457053(6)	5:76,460,705	G/A	1.08	1.09 (0.84-1.34)	0.52	G[c]	0.89 (0.66-1.21)	0.46	G	1.10 (0.92-1.28)	0.29	G[c]	0.95 (0.83-1.08)	0.43	0.52	A[c]
rs9505118(1)	6:3,672,354	A/G	1.06	1.05 (0.88-1.21)	0.59	G	0.99 (0.85-1.15)	0.85	G[c]	1.12 (0.97-1.27)	0.15	G	0.95 (0.87-1.04)	0.29	0.56	A
rs6931514(7)	6:20,811,931	G/A	1.25	1.09 (0.94-1.25)	0.27	G[c]	1.21 (1.04-1.40)	0.02[b]	G[c]	1.20 (1.03-1.38)	0.04[b]	G[c]	0.86 (0.78-0.94)	$2x10^{-3}$[b]	0.62	A[c]
rs3130501(1)	6:31,244,432	G/A	1.07	1.02 (0.84-1.19)	0.87	A	0.92 (0.78-1.08)	0.3	A[c]	0.95 (0.79-1.11)	0.51	A[c]	0.95 (0.87-1.05)	0.32	0.69	A[c]
rs9472138(7)	6:43,919,740	T/C	1.06	1.04 (0.84-1.24)	0.69	A[c]	0.87 (0.70-1.07)	0.18	A	1.18 (0.97-1.39)	0.12	A[c]	1.02 (0.91-1.16)	0.7	0.14	A[c]
rs2191349(3)	7:15,030,834	T/G	1.06	0.96 (0.78-1.13)	0.62	C[c]	0.94 (0.80-1.10)	0.44	C[c]	0.99 (0.84-1.14)	0.91	C[c]	1.04 (0.94-1.14)	0.44	0.9	A[c]
rs864745(7)	7:28,147,081	T/C	1.1	0.90 (0.72-1.07)	0.23	G[c]	0.99 (0.83-1.19)	0.91	G[c]	0.88 (0.70-1.05)	0.15	G[c]	1.09 (0.98-1.20)	0.12	0.61	A[c]
rs4607517(3)[a]	7:44,202,193	A/G	1.15	1.25 (1.01-1.49)	0.06	A[c]	1.58 (1.29-1.94)	$1x10^{-5}$[b]	A[c]	1.01 (0.78-1.25)	0.91	A[c]	1.30 (1.15-1.48)	$6x10^{-5}$	0.02	A[c]
rs7636(9)	7:100,328,013	A/G	1.85	0.98 (0.36-1.61)	0.96	A		N/A		1.22 (0.88-1.56)	0.26	A[c]	1.16 (0.85-1.58)	0.34	0.56	A[c]
rs10229583(2)	7:127,034,139	G/A	1.14	0.95 (0.76-1.14)	0.59	A[c]	1.07 (0.87-1.30)	0.54	A	1.10 (0.94-1.26)	0.23	A	1.04 (0.94-1.16)	0.41	0.51	A
rs972283(6)	7:130,117,394	G/A	1.07	0.93 (0.77-1.08)	0.34	A[c]	1.11 (0.95-1.31)	0.2	A	0.95 (0.80-1.10)	0.5	A[c]	0.99 (0.91-1.09)	0.92	0.21	A[c]
rs516946(5)	8:41,638,405	C/T	1.09	0.70 (0.46-0.94)	$4x10^{-3}$[b]	A[c]	0.91 (0.73-1.13)	0.39	A[c]	1.06 (0.85-1.27)	0.57	A	0.90 (0.79-1.02)	0.1	0.04	A[c]
rs896854(6)	8:96,029,687	T/C	1.06	1.06 (0.89-1.23)	0.52	A[c]	1.09 (0.91-1.29)	0.36	A[c]	1.03 (0.88-1.18)	0.71	A[c]	1.05 (0.96-1.16)	0.29	0.91	A[c]
rs3802177(6)	8:118,254,206	G/A	1.15	0.95 (0.80-1.11)	0.54	A[c]	0.78 (0.67-0.91)	$2x10^{-3}$[b]	A[‡]	0.92 (0.75-1.10)	0.36	A[c]	0.88 (0.80-0.96)	$7x10^{-3}$[b]	0.17	A[c]

rs7041847(4)	9:4,277,466	A/G	1.1	0.87 (0.71-1.03)	0.09	A	1.02 (0.88-1.19)	0.77	A[c]	0.90 (0.75-1.05)	0.16	G[c]	1.01 (0.92-1.10)	0.95	0.1	A[c]
rs10965250(6)[a]	9:22,123,284	G/A	1.2	0.82 (0.66-0.98)	0.02[b]	A[c]	0.79 (0.68-0.92)	$3x10^{-3b}$	A[c]	0.81 (0.61-1.01)	0.04[b]	A[c]	0.81 (0.73-0.89)	$3x10^{-5}$	0.96	A[c]
rs2796441(5)	9:83,498,768	G/A	1.07	1.08 (0.92-1.23)	0.35	G[c]	1.13 (0.97-1.33)	0.12	G[c]	0.90 (0.75-1.05)	0.17	A[c]	0.90 (0.83-0.99)	0.03 [b]	0.91	A[c]
rs11257655(10)	10:12,347,900	T/C	1.15	1.16 (1.00-1.31)	0.06	A[c]	0.92 (0.79-1.08)	0.3	G[c]	0.91 (0.74-1.08)	0.28	A	1.05 (0.96-1.15)	0.29	0.12	A[c]
rs1802295(11)	10:70,601,480	A/G	1.08	0.89 (0.67-1.11)	0.29	A	0.98 (0.77-1.26)	0.89	A	0.90 (0.72-1.09)	0.28	A	0.92 (0.81-1.04)	0.17	0.84	A
rs12571751(5)	10:80,612,637	A/G	1.08	0.88 (0.72-1.04)	0.12	G[c]	0.96 (0.82-1.12)	0.62	G[c]	0.90 (0.76-1.05)	0.18	G[c]	1.09 (1.00-1.20)	0.06	0.74	A[c]
rs1111875(12)	10:94,337,810	G/A	1.18	1.14 (0.97-1.30)	0.12	G[c]	1.22 (1.02-1.45)	0.03[b]	G[c]	1.12 (0.96-1.27)	0.16	G[c]	0.89 (0.80-0.98)	1x10[-3b]	0.46	A[c]
rs6583826(9)	10:94,452,862	C/T	1.13	1.17 (1.00-1.34)	0.06	G[c]	1.28 (1.09-1.52)	$3x10^{-3b}$	G[c]	1.01 (0.84-1.17)	0.95	G[c]	0.85 (0.78-0.94)	0.02 [b]	0.24	A[c]
rs10885122(3)	10:113,032,083	G/T	1.04	1.27 (1.00-1.54)	0.08	A	0.94 (0.70-1.26)	0.66	A[c]	0.97 (1.80-1.13)	0.68	A[c]	1.02 (0.90-1.16)	0.77	0.22	A
rs7903146(6)[a]	10:114,748,339	T/C	1.4	1.21 (0.93-1.49)	0.19	A[c]	1.08 (0.70-1.68)	0.72	A[c]	1.37 (1.21-1.53)	$2x10^{-4b}$	A[c]	1.30 (1.13-1.49)	$2x10^{-4}$	0.52	A[c]
rs231362(6)	11:2,648,047	G/A	1.08	1.24 (1.03-1.45)	0.04[b]	A	0.95 (0.71-1.27)	0.73	A[c]	0.88 (0.70-1.05)	0.14	A[c]	1.01 (0.88-1.13)	0.99	0.05	A
rs5215(13)	11:17,365,206	C/T	1.14	1.05 (0.90-1.21)	0.52	G[c]	0.95 (0.81-1.11)	0.5	G	1.12 (0.97-1.28)	0.15	G[c]	0.96 (0.88-1.06)	0.41	0.34	A[c]
rs7944584(3)	11:47,292,896	A/T	1.01	0.91 (0.52-1.29)	0.62	T[c]	0.60 (0.37-1.00)	0.05	T[c]	0.94 (0.75-1.13)	0.52	T[c]	1.12 (0.95-1.32)	0.17	0.29	A[c]
rs174550(3)	11:61,328,054	T/C	1.04	1.18 (1.00-1.35)	0.07	A[c]	1.27 (1.08-1.48)	$4x10^{-3b}$	A[c]	0.94 (0.71-1.18)	0.63	G[c]	1.19 (1.07-1.33)	1x10[-3 b]	0.45	A[c]
rs1552224(6)	11:72,110,746	A/C	1.14	0.63 (0.33-0.92)	$2x10^{-3b}$	C[c]	1.04 (0.78-1.41)	0.78	C	0.84 (0.65-1.03)	0.07	C[c]	1.21 (1.04-1.39)	0.01 [b]	0.06	A[c]
rs1387153(6)	11:92,313,476	T/C	1.09	0.93 (0.78-1.08)	0.36	G[c]	1.12 (0.96-1.31)	0.14	A[c]	1.20 (1.04-1.35)	0.02[b]	G	1.08 (0.99-1.18)	0.1	0.07	A[c]
rs10842994(5)	12:27,856,417	C/T	1.1	0.86 (0.67-1.04)	0.1	A[c]	0.98 (0.81-1.19)	0.83	A[c]	0.87 (0.66-1.08)	0.18	A[c]	0.90 (0.80-1.01)	0.07	0.58	A[c]

rs35767(3)	12:101,399,699	G/A	1.04	0.99 (0.82-1.15)	0.88	A[c]	1.06 (0.91-1.24)	0.47	A	0.76 (0.58-0.94)	$3x10^{-3}$[b]	A[c]	0.94 (0.86-1.04)	0.25	0.02	A[c]
rs4275659(1)	12:119,887,315	C/T	1.14	1.04 (0.88-1.20)	0.65	A	1.01 (0.87-1.18)	0.86	G[c]	0.92 (0.77-1.07)	0.26	A[c]	1.01 (0.92-1.10)	0.49	0.56	A
rs7305618(14)	12:122,013,881	C/T	1.06	0.97 (0.82-1.13)	0.74	A[c]	0.96 (0.81-1.13)	0.61	A[c]	1.01 (0.85-1.17)	0.94	A	0.97 (0.88-1.06)	0.87	0.95	A[c]
rs1359790(15)	13:79,615,157	G/A	1.15	0.80 (0.58-1.02)	0.04[†]	A[c]	0.95 (0.80-1.13)	0.56	A[c]	1.03 (0.84-1.23)	0.74	A	0.93 (0.83-1.05)	0.24	0.24	A[c]
rs1436953(16)	15:60,201,306	G/A	1.14	0.85 (0.70-1.01)	0.05	A[c]	1.06 (0.90-1.26)	0.49	A	1.00 (0.85-1.15)	0.99	A	0.97 (0.88-1.06)	0.47	0.18	A[c]
rs7177055(5)	15:75,619,817	A/G	1.08	1.21 (1.05-1.37)	0.02[b]	A[c]	1.08 (0.92-1.27)	0.33	A[c]	1.05 (0.90-1.20)	0.5	A[c]	1.11 (1.01-1.21)	0.02[b]	0.43	A[c]
rs11634397(6)	15:78,219,277	G/A	1.06	0.98 (0.80-1.17)	0.87	G	0.99 (0.75-1.30)	0.92	G	0.93 (0.78-1.08)	0.33	A[c]	0.97 (0.87-1.08)	0.56	0.74	A[c]
rs8042680(6)	15:89,322,341	A/C	1.07	1.17 (0.84-1.50)	0.35	C	N/A			1.18 (1.00-1.36)	0.06	C	0.85 (0.72-0.99)	0.04[b]	0.96	A
rs9939609(17)[a]	16:52,378,028	A/T	1.25	1.32 (1.15-1.49)	$1x10^{-3}$[b]	A[c]	1.16 (0.93-1.45)	0.18	A[c]	1.16 (1.01-1.32)	0.06	A[c]	1.22 (1.10-1.35)	$2x10^{-4}$	0.51	A[c]
rs7202877(5)	16:73,804,746	T/G	1.12	1.01 (0.80-1.22)	0.89	C	0.90 (0.75-1.09)	0.29	C[c]	1.07 (0.82-1.32)	0.59	C	1.03 (0.91-1.16)	0.66	0.51	A[c]
rs391300(18)	17:2,163,008	G/A	1.28	0.92 (0.77-1.07)	0.29	A[c]	0.91 (0.77-1.06)	0.23	A[c]	1.04 (0.89-1.20)	0.59	G[c]	0.93 (0.85-1.02)	0.11	0.88	A[c]
rs12970134(5)[a]	18:56,035,730	A/G	1.08	1.19 (0.97-1.41)	0.12	A[c]	1.17 (0.96-1.44)	0.13	A[c]	1.28 (1.12-1.44)	$3x10^{-3}$[b]	A[c]	1.22 (1.10-1.36)	$3x10^{-4}$	0.77	A[c]
rs3786897(4)	19:38,584,848	A/G	1.1	0.95 (0.80-1.11)	0.56	G[c]	0.91 (0.78-1.07)	0.24	G[c]	1.05 (0.88-1.22)	0.59	G	1.04 (0.94-1.14)	0.47	0.51	A[c]
rs4812829(11)	20:42,422,681	A/G	1.09	1.03 (0.87-1.18)	0.75	A[c]	1.11 (0.95-1.29)	0.2	A[c]	1.04 (0.89-1.20)	0.61	A[c]	1.06 (0.97-1.16)	0.23	0.78	A[c]

[1] SNP previously associated with type 2 diabetes or fasting plasma glucose at genome-wide significance ($P<5x10^{-8}$), with original reference. [2] Risk allele from previous study. [3] Other allele from previous study. [4] Odds ratio from previous study. [5] Effect allele from analysis in Malaysian sample.

[a] Denotes significance at $P<8.1x10^{-4}$

[b] Denotes significance at $P<0.05$

[c] Denotes same effect direction as previously reported

N/A denotes missing data to MAF <0.01

Supplementary Table 2: SNPs reaching $P<1x10^{-5}$ from the Metabochip-wide meta-analysis

SNP	Chr:BP	Context	Gene	A1*	A2	Odds Ratio	95% CI	*P*
chr16:52362235**	chr16:52362235	Intronic	*FTO*	A	G	0.15	0.07, 0.33	$3.41x10^{-6}$
rs61316436	chr12:101257553	Intergenic	–	A	G	4.28	2.31, 7.93	$3.97x10^{-6}$
rs576674	chr13:32452302	Intergenic	–	A	G	0.78	0.7, 0.87	$6.95x10^{-6}$
rs740093	chr7:44636189	Intronic	*OGDH*	A	G	0.81	0.74, 0.89	$7.28x10^{-6}$
rs217378	chr7:44574628	Intronic	*DDX56*	A	G	0.81	0.74, 0.89	$8.52x10^{-6}$

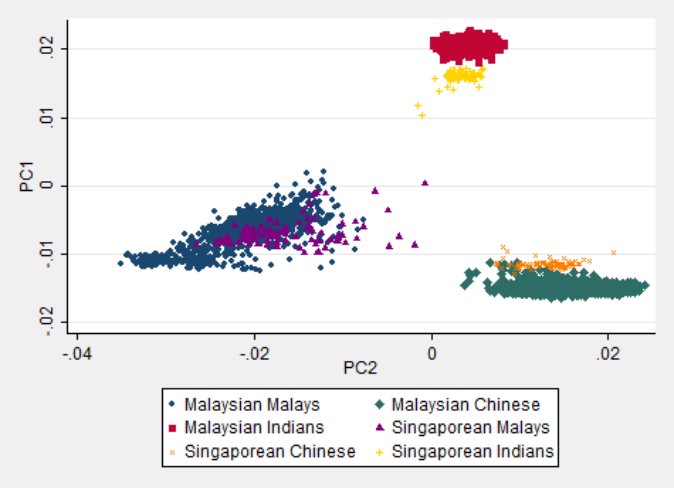
Notes * Modelled allele corresponding to the OR. ** Metabochip variant with no corresponding rs number.

References

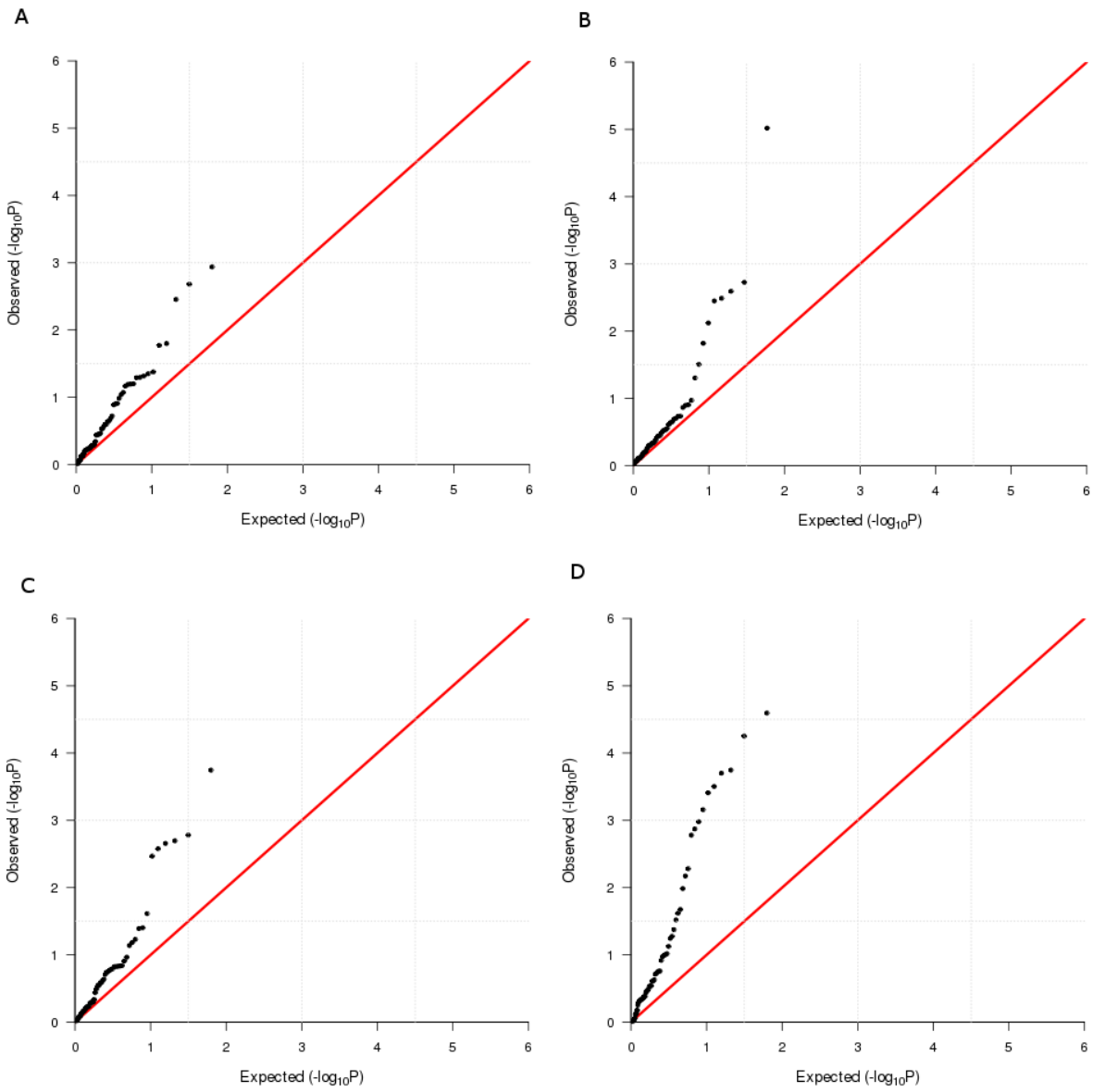
1. Replication DIG, Meta-analysis C. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nature genetics. 2014;46(3):234-44.
2. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64-91.
3. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics. 2010;42(2):105-16.
4. 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. 2012;44(1):67-72.
5. 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(9):981-90.
6. 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(7):579-89.
7. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature genetics. 2008;40(5):638-45.
8. 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(5829):1341-5.
9. 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(4):e1001363.
10. Li H, Gan W, Lu L, Dong X, Han X, Hu C, et al. A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. Diabetes. 2013;62(1):291-8.
11. 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(10):984-9.
12. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature genetics. 2012;44(9):991-1005.
13. 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(5829):1336-41.
14. 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(8):2038-46.
15. 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(9).
16. Cui B, Zhu X, Xu M, Guo T, Zhu D, Chen G, et al. A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese. PloS one. 2011;6(7):e22353.
17. 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 genetics. 2012;8(5):e1002741.

18. 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(2):e1000847.

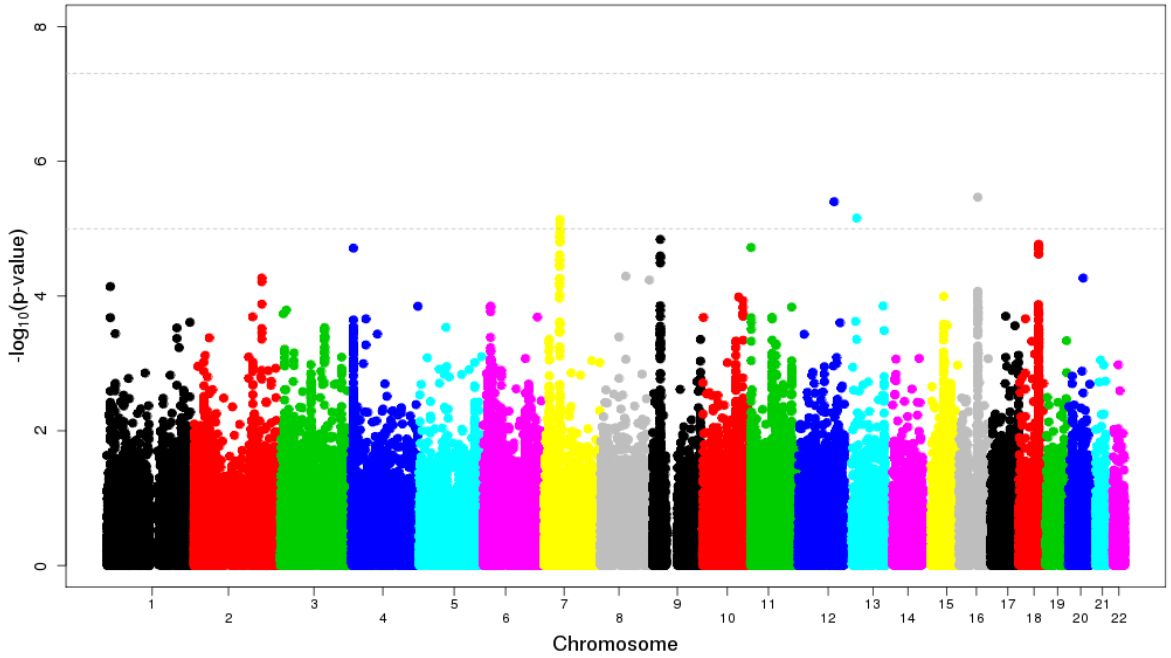
Data Supplement



Supplementary Figure 1: Plot of the first two ancestry principal components in Malaysian samples, incorporating reference data from the Singaporean Genome Variation Project.



Supplementary Figure 2: Quantile-quantile (Q-Q) plots of the *P*-values for association of 62 SNPs observed in each of ethnic groups and meta-analysis. (A) Malays, (B) Chinese, (C) Indians, (D) Combined meta-analysis



Supplementary Figure 3: Manhattan plot of Metabochip-wide meta-analysis results for all SNPs passing quality control in at least two of the three ethnic groups. The genomic inflation factor (λ_{GC}) was 1.054.

References

1. Replication DIG, Meta-analysis C. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nature genetics. 2014;46(3):234-44.
2. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64-91.
3. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics. 2010;42(2):105-16.
4. 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. 2012;44(1):67-72.
5. 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(9):981-90.
6. 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(7):579-89.
7. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature genetics. 2008;40(5):638-45.
8. 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(5829):1341-5.
9. 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(4):e1001363.
10. Li H, Gan W, Lu L, Dong X, Han X, Hu C, et al. A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. Diabetes. 2013;62(1):291-8.
11. 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(10):984-9.
12. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature genetics. 2012;44(9):991-1005.
13. 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(5829):1336-41.
14. 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(8):2038-46.
15. 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(9).
16. Cui B, Zhu X, Xu M, Guo T, Zhu D, Chen G, et al. A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese. PloS one. 2011;6(7):e22353.
17. 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 genetics. 2012;8(5):e1002741.

18. 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(2):e1000847.

References

1. Replication DIG, Meta-analysis C. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nature genetics. 2014;46(3):234-44.
2. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64-91.
3. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics. 2010;42(2):105-16.
4. 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. 2012;44(1):67-72.
5. 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(9):981-90.
6. 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(7):579-89.
7. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature genetics. 2008;40(5):638-45.
8. 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(5829):1341-5.
9. 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(4):e1001363.
10. Li H, Gan W, Lu L, Dong X, Han X, Hu C, et al. A genome-wide association study identifies GRK5 and RASGRP1 as type 2 diabetes loci in Chinese Hans. Diabetes. 2013;62(1):291-8.
11. 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(10):984-9.
12. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature genetics. 2012;44(9):991-1005.
13. 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(5829):1336-41.
14. 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(8):2038-46.
15. 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(9).
16. Cui B, Zhu X, Xu M, Guo T, Zhu D, Chen G, et al. A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese. PloS one. 2011;6(7):e22353.
17. 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 genetics. 2012;8(5):e1002741.

18. 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(2):e1000847.

CHAPTER 4: QUANTIFYING THE ROLES OF CLASSICAL RISK FACTORS IN TYPE 2 DIABETES USING A MULTI-ETHNIC MALAYSIAN COHORT

Abdullah N, Abdul Murad N. A, Attia J, Oldmeadow C, Kamaruddin M.A., Abd. Jalal N., Ismail N., Jamal R, Scott R.J, Holliday E. G.

Submitted

4.1 Statement of Co-authors

"As co-authors of the paper:
Abdullah N, Abdul Murad N. A, Attia J, Oldmeadow C, Kamaruddin M.A., Abd. Jalal N., Ismail N., Jamal R, Scott R.J, Holliday E. G. Quantifying the Roles of Classical Risk Factors in Type 2 Diabetes using a Multi-ethnic Malaysian Cohort. Submitted, we confirm that Noraidatulakma Abdullah contributed to this publication by performing the analysis, interpreting the result and writing the manuscript."

Co-Authors:

Dr Elizabeth Holliday
Signed
Date: 31 July 2017

Professor John Attia
Signed
Date: 1st August 2017

Dr Christopher Oldmeadow
Signed
Date: 1st August 2017

Professor Rodney Scott
Signed
Date: 2/8/2017

Professor Datuk A. Rahman A. Jamal
Signed
Date:

Dr Nor Azian Abdul Murad
Signed
Date:

Mohd Arman Kamaruddin
Signed
Date:

Nazihah Abd Jalal

Signed

Date:

Norliza Ismail

Signed

Date:

Faculty Assistant Dean (Research Training)

Professor Robert Callister

Signed

Date: 3/08/2017

4.2 Summary of Publication 3

Introduction

The T2D epidemic is growing rapidly in Asian countries in conjunction with rapid urbanisation and modernisation, leading to transitions to more “Western” diet and lifestyle including reductions in physical activity. It has previously been shown in other populations that only a small proportion of T2D heritability can be explained by known, common T2D genetic variant. Findings from the genetic study reported in Chapter 2 suggested a similar result applies in the Malaysian population, with a genetic risk score aggregating 62 known T2D variants explaining only about 2% of overall T2D risk variation in any ancestral group. Thus, the escalating prevalence of T2D in Malaysia appears unlikely to solely reflect the effects of common genetic variants. Rather, lifestyle changes may have a more substantial influence. In this chapter, the association between environmental/lifestyle risk factors and T2D were assessed in the Malaysian population. A predictive model was constructed to assess the combined contribution of lifestyle risk factors to T2D variation in the Malaysian population, with this being estimated and compared between ancestral groups.

International Journal of Public Health

Quantifying the Roles of Classical Risk Factors in Type 2 Diabetes using a Multi-ethnic Malaysian Cohort

--Manuscript Draft--

Manuscript Number:	IJPH-D-16-00403
Full Title:	Quantifying the Roles of Classical Risk Factors in Type 2 Diabetes using a Multi-ethnic Malaysian Cohort
Article Type:	Original article
Keywords:	Type 2 Diabetes, Classical Risk Factors, Population differences, Asian Populations
Corresponding Author:	Elizabeth Holliday Hunter Medical Research Institute Newcastle, New South Wales AUSTRALIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Hunter Medical Research Institute
Corresponding Author's Secondary Institution:	
First Author:	Noraidatulakma Abdullah, Bsc(Hons), Msc
First Author Secondary Information:	
Order of Authors:	Noraidatulakma Abdullah, Bsc(Hons), Msc
	Nor Azian Abdul Murad
	John Attia
	Christopher Oldmeadow
	Mohd Arman Kamaruddin
	Nazihah Abd Jalal
	Norliza Ismail
	Rahman Jamal
	Rodney Scott
	Elizabeth Holliday
Order of Authors Secondary Information:	
Funding Information:	The National Biotechnology Division, Ministry of Science, Technology and Innovation (MOSTI), Malaysia (ER-05-01-02-MEB001) — Prof Rahman Jamal
	Ministry of Higher Education of Malaysia and Universiti Kebangsaan Malaysia — Mrs Noraidatulakma Abdullah
Abstract:	Objectives We assessed the association and predictive utility of classical risk factors for type 2 diabetes in Malay, Chinese and Indian participants from The Malaysian Cohort project. The objective was to assess whether recent increases in type 2 diabetes prevalence may be predominantly due to environmental factors. Methods This nested case-control study involved 4077 samples of Malay, Chinese and Indian. Association of lifestyle-related risk factors was assessed via logistic regression. The

	combined contribution of the factors to type 2 diabetes risk was compared using Area Under the Receiver-Operating Characteristic curve and 95% confidence interval. Effect modification by ancestry was assessed for associated risk factors. Results Age and waist-to-hip ratio were significantly associated with type 2 diabetes. In models including age, gender, waist-to-hip ratio and physical activity, the AUC ranged from 0.75 to 0.83, being significantly higher in Chinese than Malays or Indians. Conclusion This study suggests that obesity and advancing age are major drivers of the escalating Malaysian type 2 diabetes prevalence. Interventions targeting these factors could reduce the burden of type 2 diabetes.
Suggested Reviewers:	Richard Jensen University of Washington richaj@uw.edu
	Mary Frances Cotch National Institutes of Health, USA mfc@nei.nih.gov
	Amanda Cox Griffith University, Australia a.cox@griffith.edu.au

Name	Institution	City	Country	Email address
Noraidatulakma Abdullah	School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle	Newcastle, New South Wales	Australia	Noraidatulakma.Abdullah@uon.edu.au
Nor Azian Abdul Murad	UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia	Kuala Lumpur	Malaysia	nor_azian@ppukm.ukm.edu.my
John Attia	Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, Faculty of Health, University of Newcastle	Newcastle, New South Wales	Australia	john.attia@newcastle.edu.au
Christopher Oldmeadow	Clinical Research Design, IT and Statistical Support (CReDITSS) Unit, Hunter Medical Research Institute	Newcastle, New South Wales	Australia	Christopher.Oldmeadow@hmri.com.au
Mohd Arman Kamaruddin	UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia	Kuala Lumpur	Malaysia	arman@ppukm.ukm.edu.my
Nazihah Abd Jalal	UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia	Kuala Lumpur	Malaysia	jiha_jalal@yahoo.com
Norliza Ismail	UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia	Kuala Lumpur	Malaysia	norliza.ismail212@gmail.com
Rahman Jamal	UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia	Kuala Lumpur	Malaysia	rahmanj@ppukm.ukm.edu.my
Rodney J. Scott	School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle	Newcastle, New South Wales	Australia	rodney.scott@newcastle.edu.au
Elizabeth G. Holliday	Clinical Research Design, IT and Statistical Support (CReDITSS) Unit, Hunter Medical Research Institute	Newcastle, New South Wales	Australia	elizabeth.holliday@hmri.org.au

Quantifying the Roles of Classical Risk Factors in Type 2 Diabetes using a Multi-ethnic Malaysian Cohort

Keyword: Type 2 Diabetes, Classical Risk Factors, Population differences, Asian Populations

Word count: 2896

No of tables: 3

No of figures: 2

No of references: 40

No of supplementary file: 1

Abstract

Objectives

We assessed the association and predictive utility of classical risk factors for type 2 diabetes in Malay, Chinese and Indian participants from The Malaysian Cohort project. The objective was to assess whether recent increases in type 2 diabetes prevalence may be predominantly due to environmental factors.

Methods

This nested case-control study involved 4077 samples of Malay, Chinese and Indian. Association of lifestyle-related risk factors was assessed via logistic regression. The combined contribution of the factors to type 2 diabetes risk was compared using Area Under the Receiver-Operating Characteristic curve and 95% confidence interval. Effect modification by ancestry was assessed for associated risk factors.

Results

Age and waist-to-hip ratio were significantly associated with type 2 diabetes. In models including age, gender, waist-to-hip ratio and physical activity, the AUC ranged from 0.75 to 0.83, being significantly higher in Chinese than Malays or Indians.

Conclusion

This study suggests that obesity and advancing age are major drivers of the escalating Malaysian type 2 diabetes prevalence. Interventions targeting these factors could reduce the burden of type 2 diabetes.

Introduction

There are 387 million people living with diabetes worldwide and about 90 percent of these have type 2 diabetes (T2D). The prevalence of T2D is growing rapidly, especially in Asian countries, with about 60% of diabetes patients worldwide currently living in Asia (Hu 2011).

Asians have been found to be at increased risk of developing T2D compared to people of European ancestry (Chan et al. 2009). While the causes are not entirely clear, recent demographic and lifestyle transitions are likely important. T2D prevalence in Asia has increased in parallel with rapid economic development, an ageing population, urbanisation, nutritional transitions, reduced physical activity, and changes in other lifestyle patterns (Abdullah N. 2014). Adoption of westernised dietary patterns has been associated with T2D and partly reflects increased caloric intake from animal fat, which has almost doubled in India and China in recent decades (Misra et al. 2011; Wang et al. 2007). Asian populations also have the highest correlation between per capita sugar consumption and T2D prevalence among 165 countries (Weeratunga et al. 2014). Compounding the change in diet, physical activity has also significantly reduced in Asian populations due to rapid urbanization and modernisation (Ng et al. 2009).

In contrast, the contribution of known genetic variants to disease risk appears to be small in Asian populations. Recent genetic studies conducted among East Asian ancestry suggests that known T2D genetic risk variants explain only about 2% of variation in disease risk (Chen et al. 2014).

Among Asian countries, Malaysia has one of the highest comparative prevalence of T2D (International Diabetes Federation 2013), although the country has been relatively understudied in T2D research. Malaysia has a total population of 28.3 million (Department of Statistics Malaysia 2011) and its multi-ethnic society includes three major ancestral groups: Malays (~63%); Chinese (~25%); and Indian (~7%). The prevalence of T2D appears to differ among these three groups, with Malaysian Indians having the highest prevalence (25% to 28%), followed by Malays (17% to 19%) and Chinese (9% to 14%) (Jamal R. 2014).

Similar to other Asian countries, the rising T2D prevalence in Malaysia appears unlikely to reflect the effects of known, common genetic risk variants. In a recent study, a genetic risk score aggregating 62 validated T2D genetic risk variants explained less than 2% of overall T2D risk in any of the three major ancestral groups (Abdullah et al. 2015). As in other Asian countries, lifestyle factors may be more important and quantifying their effect on population disease risk may help to identify targets for public health intervention. With this

intent, we assessed the association and predictive utility of classical T2D risk factors in samples from the three major Malaysian ancestral groups. This study had three principal aims: i) to assess the association between known and novel lifestyle risk factors and T2D, both within and across Malay, Chinese and Indian Malaysian populations, ii) to estimate and compare the combined contribution of all disease-associated risk factors to T2D in and between three ancestral groups, and iii) via interaction analyses, to identify risk factors whose effect on T2D may be modified by ancestry.

Methods

Data sources and study samples

This study was a nested case-control study using participants from the Malaysian Cohort project (MCP), a prospective multi-ethnic, population-based cohort including 106,527 volunteers aged between 35 and 70 years (Jamal R. 2014). For this analysis, we randomly selected approximately 600-700 samples from each major ancestral group (Malay, Chinese and Indian) with fasting plasma glucose (FPG) exceeding 7.5 mmol/L (or 126 mg/dL) as T2D cases, and a similar number of ancestry-matched controls with FPG <5.5 mmol/L (or 99 mg/dL). Ethnicity was defined using the self- reported ethnicity of the participant and their family for three preceding generations.

A total of 4077 samples were used in this analysis: 1,323 Malays (600 cases, 723 controls), 1344 Chinese (654 cases, 690 controls) and 1,410 Indians (708 cases, 702 controls). All relevant approvals for The Malaysia Cohort were granted by the institutional review and ethics board of the Universiti Kebangsaan Malaysia, in accordance with the declaration of Helsinki. All participants gave written, informed consent.

Risk factor selection

We selected known T2D risk factors using evidence from previous studies (Cahill et al. 2014). These comprised age, gender, physical activity, body mass index (BMI), waist circumference (WC), waist-to-hip ratio (WHR), deep fried food consumption and coffee consumption. We also selected two potentially relevant, novel risk factors: sautéed food consumption and coconut milk intake, based on evidence for foods high in trans-fat being associated with cardiometabolic disease and insulin resistance (Lopez-Garcia et al. 2005).

Questionnaire-derived variables

Information related to demographic and environment factors was collected by questionnaires and interviews at baseline (Jamal R. 2014). To assist coefficient interpretation, we categorised age into 3 groups: < 50 years; 50-60 years, and >60 years. Dietary variables were measured by asking participants how often they had consumed foods of a particular type or prepared using specific methods in the preceding week. Questions had five response choices which were categorised into 3 groups: less than once a week, 1 to 3 times a week and more than 4 times a week following a previous study (Qi et al. 2014).

Self-reported physical activity was classified using self-reported average weekly vigorous activity over the last four months (see Online Resource1 for list of activities), which we categorised as either active or inactive using a threshold of 150 minutes per week (World Health Organization 2011).

Anthropometric measurements

Height, weight, body mass index (BMI), waist circumference (WC) and waist- to-hip ratio (WHR) were measured three times and averaged. BMI was categorised as: <25 kg/m^2 (normal), 25-30 kg/m^2 (pre-obese) and >30 kg/m^2 (obese). For WC and WHR, sex-specific cut-offs were used to derive three categories (World Health Organization 2008). For WC these were: low risk (males: < 94cm; females: <80cm); moderate risk (males: 94-102cm; females: 80-88cm) and; high risk (males: >102 cm; females: >88cm). WHR was categorised as: low risk (< males: <0.95; females: <0.80); moderate risk (males: 0.96-1.0; females: 0.81-0.85) and; high risk (males: >1; females: >0.85).

Missing data handling by multiple imputation

Both complete case and multiple imputation analyses were performed. Multiple imputations were performed by chained equations (MICE) with 25 cycles. In each cycle, missing values in each variable were imputed based on a predictive distribution derived from regression on all other variables in the imputation model (gender, age group, waist-to-hip ratio and physical activity).

Statistical Analyses

We used multivariable logistic regression modelling to investigate associations between the lifestyle factors of interest and type 2 diabetes within each of the three ancestral groups separately and in the combined population. For each analysis we used a variable selection process as previously described (Greenland S. Pearce N. 2014). We initially fitted a multivariable model including all selected risk factors then removed the least significant risk factor (P>0.20) one at a time provided the likelihood ratio P-value exceeded 0.20 and the estimated coefficients (on the logit scale) of the remaining variables did not differ by more than 10%. To ensure final models were

comparable across ancestral groups, we re-included in the final model for each group any risk factor that had been removed for that group but retained for any other group. In all analyse, WHR was treated as both a continuous and categorical (but omitting gender from analysis because WHR used gender-specific cut-points). The model for the three ancestral groups combined, included ethnicity as a fixed effect. The risk explained by the classical risk factors was estimated using McFadden's pseudo R^2. Based on the final model for each ancestral group, we calculated the Area Under the Receiver-Operating Characteristic (AUROC) curve and its 95% confidence interval.

Based on the available sample size, in each ethnic group we had 80% power at α=0.05 to detect an odds ratio of 1.36 for a risk factor with prevalence in controls of 0.35. For prevalences of 0.25 and 0.15, we could detect odds ratios of 1.40 and 1.48, respectively, in each ethnic group.

We assessed effect modification of risk factor effects by ancestral group on both additive and multiplicative scales. Results were presented based on recommendations by Knol and VanderWeele (Knol and VanderWeele 2012). Effect-modification analyses were performed only using complete-case data (Knol and VanderWeele 2012). Effect modification on the additive scale was assessed by calculating the "relative excess risk due to interaction" (*RERI*). All analyses were performed using STATA 11.2 (Stata Corporation, College Station, Texas).

Results

Baseline characteristics for all available data from the full sample (N=4077) are shown in Table 1. Baseline characteristic for participants used in complete-case analyses (N=2478) are shown in Online Resource 2. Missing data were substantively due to the physical activity variable, resulting from a transition between two versions of physical activity questionnaires during the study. Descriptive statistics showed that cases were generally more likely to be male, older, obese, have higher waist-to-hip ratio and larger waist circumference than controls across all ancestry groups. Dietary variables showed similar distributions between cases and controls.

The final multivariable model included gender, age, waist-to-hip ratio and physical activity (Table 2). WC was removed due to high collinearity with WHR, and because WHR has been found to be a superior predictor of diabetes risk (Kaur et al. 2008; Xin et al. 2012). BMI did not explain significant variation in the outcome when WHR was included in the model so was also removed. None of the other risk factors (deep fried food consumption frequency, frequency of drinking coffee, consumption of sautéed food or coconut milk intake) showed association with T2D (at $P<0.2$) or were retained in the multivariable model.

In the primary analysis, the association of gender was not statistically significant (at $P<0.05$) in any ancestral group, although females appeared to have a tendency for higher disease risk across groups (Figure 1). Older people had significantly higher risk of T2D than younger people in all three ancestral groups. Compared to participants aged <50, those aged 50 to 60 years had odds ratios of 2.61 in Malays (95%CI: 1.88, 3.65), 2.32 in Indians (95%CI: 1.70, 3.16) and 1.34 in Chinese (95%CI: 0.92, 1.96). For participants aged >60, the highest risk estimate was in Chinese (OR: 3.57; 95% CI: 2.08, 6.14) followed by Indians (OR: 2.13; 95%CI: 1.24, 3.64) and Malays (OR: 1.99; 95% CI: 1.07, 3.70), although all confidence intervals overlapped. Larger WHR ratio was significantly associated with increased risk of T2D in all ancestry groups (Chinese OR: 5.81, 95%CI: 4.19, 8.05; Indians OR: 3.60, 95%CI: 2.80, 4.63; Malays OR: 3.11, 95% CI: 2.39, 4.04). Physical inactivity was not significantly associated with disease in any group, although point estimates of odds ratios were all >1.

In the results of all groups combined, results were similar, with risk of T2D increasing significantly with advancing age and increasing WHR (Table 2). The alternative model based on a categorical WHR yielded analogous results, (see Online Resource 3). Results from multiply imputed data (N=4077) were similar to the complete case analysis (N=2478) but had smaller standard errors and tighter confidence intervals (Table 2).

Based on the pseudo-R^2, the combination of gender, age, WHR and physical inactivity explained about 15.1% disease risk in Malays, 26.3% in Chinese and 18.5% in Indians (Table 2). In the combined sample, these four factors explained an estimated 19.5% of T2D risk. Results from multiply imputed data were similar (Table 2).

The area under the receiver operating characteristic curve (AUC) for the final model was highest in Chinese (AUC: 0.83, 95% CI: 0.80, 0.86), followed by Indians (AUC: 0.78, 95% CI: 0.75, 0.81) and Malays (AUC: 0.75, 95% CI: 0.72, 0.78) (Figure 2). In the combined group, the AUC was 0.79 (0.77, 0.81) (Figure 2). A test of equality of three AUC estimates (Cleves. 2002) showed a globally significant difference (P<0.001) between the three ethnic groups (Table 3). In pairwise comparisons, the Chinese AUC was significantly different to both the Malay (P<0.001) and Indian AUC (P=0.02). There was no difference between estimates in Malays and Indians (P=0.21) (Table 3).

Nominally significant (P<0.05) multiplicative interaction was observed between ancestry and a high risk waist-to-hip ratio value (compared to low risk, see Online Resource 4), with the effect of high risk waist-to-hip ratio being greater in Chinese than the reference of Malays (OR=1.81, P=0.02), and with Indians having a similar effect to Malays (OR=1.09, P=0.72). Alternatively, the effect of intermediate age (50-60, compared to <50) was significantly lower in Chinese than Malays (Online Resource 5), on both multiplicative (OR=0.55, P=0.01) and additive (OR=0.22, P=0.002) scales. The effect in Indians was non-significantly lower than Malays (OR=0.90 and 0.47 on multiplicative and additive scales, respectively). There was no evidence of interaction between ancestral group and physical activity (Online Resource 6).

Discussion

Our findings suggested that four risk factors account for about 20% of case control variation in T2D in the Malaysian population: age, gender, WHR and physical inactivity. Waist-to-hip ratio and age consistently showed significant association with disease across ancestral groups. This suggests that the major contributors to the increasing T2D prevalence in Malaysia are determinants of obesity such as diet and physical inactivity, together with the ageing population.

Abdominal obesity increases the risk of T2D by increasing the secretion of non-esterified fatty acids and adipocytokines such as tumour necrosis factor-α and reducing adiponectin, leading to insulin resistance and T2D (Despres and Lemieux 2006). Asian individuals have been found to have a higher distribution of body fat around organs and in the abdominal area with concomitantly lower muscle mass, compared to Europeans with the same healthy BMI or WC (Lear et al. 2007). Within Asian groups, a previous study found that body fat percentage tends to be naturally higher in Indians, followed by Malays and Chinese (Deurenberg et al. 2002). In our sample, Chinese controls also had a lower level of adiposity than Malays or Indians, corresponding with their lower overall T2D prevalence. However, the risk of T2D resulting from increasing adiposity was greater in Chinese compared than the other ancestral groups. An analogous result was observed in Chinese participants in the Multi-Ethnic Study of Atherosclerosis (Lutsey et al. 2010). These findings have public health significance, suggesting a greater risk of diabetes resulting from obesity in Chinese individuals. The findings may also reflect anthropometric differences between ancestral groups, suggesting the possible utility of ethnicity-specific anthropometric cut- points for estimating diabetes risk.

The odds of T2D was also observed to increase with advancing age (Alberti et al. 2007). This is known to be due to age-related reductions in skeletal muscle mass (sarcopenia) and activation of glycogen synthase, and increases in visceral adiposity, leading to insulin resistance and glucose intolerance (Landi et al. 2013).

High intake of trans-fat measured by deep frying, sautéing and use of coconut milk has been associated with increased cardiometabolic risk and insulin resistance (Lopez-Garcia et al. 2005) while coffee consumption is positively related to insulin sensitivity and improves pancreatic beta-cell function (Loopstra-Masters et al. 2011). However, none of these dietary factors showed association with T2D in our Malaysian sample. While measurement error is a possible explanation, other dietary factors have a greater impact onT2D risk. Dietary factors not assessed in this study but which have previously shown association with T2D include polished rice and refined wheat, which are staple foods in Asia (Villegas et al. 2007) and sugar-sweetened beverages (Malik

et al. 2006) and Western-style fast food (Odegaard et al. 2012) consumption of which have increased in recent years. These may have a larger influence on T2D risk in Malaysian populations.

Although physical activity showed a positive effect, the association was not statistically significant. Self-report questionnaires can bias measurement of physical activity as they are highly dependent on participants' memory and suffer from recall and reporting bias (Prince et al. 2008). Objective measurements of physical activity, e.g., using a pedometer or accelerometer provide more reliable results than self-report (Kelly et al. 2011). A recent study found poor correlation between self-reported physical activity and pedometer-assessed step count (Ewald et al. 2010). These limitations may have caused over- or underreporting of physical activity in this study.

Another potential limitation was that this was a cross-sectional study, and thus unable to assign a temporal sequence to a cause and its effect. For this reason, reverse-causation is possible, whereby the association of T2D with measured exposure could in part reflect bias attributable to the outcome's effect on measured exposure.

The predictive ability of our four-factor models across ancestral groups ranged from 0.75 to 0.8, indicating high reliability and good capacity to discriminate between persons who developed T2D from those that did not. Although there have been numerous predictive models constructed for T2D, (Ye et al. 2014) there have been limited data comparing different Asian populations, and few studies assessing Malays. Our AUROC estimates were higher in Chinese than Malays and Indians. This substantially reflects the relatively larger effect of WHR in Chinese, as discussed above. This larger effect was also observed as a significant interaction between larger WHR and Chinese ancestry, suggesting that Individuals of Chinese ancestry with higher WHR ratio may be an important subpopulation for targeted T2D interventions.

Although our models performed well, they still leave a substantial proportion of T2D risk unexplained. Further, although our study showed that Chinese have higher odds of disease based on classical risk factors, the prevalence of T2D in Malaysia is highest in Indians. Ancestral differences in T2D prevalence may partly reflect ancestry-specific interactions between genetic and environmental factors (Wulan et al. 2010) and gene-environment interaction studies may provide insights into T2D risk differences in the multiethnic Malaysian population. Another possible contributor is epigenetic modification, (Ling and Groop 2009) with studies reporting that DNA methylation is influenced by diet and exercise (Ronn et al. 2013) and that methylation scores at T2D risk loci differ between Asians and Europeans (Chambers et al. 2015).

Aside from these limitations, our study has identified T2D risk factors shared across multiethnic populations in Malaysia and differing in relative effect between individual populations. These results help to identify susceptible population subgroups that may benefit from targeted intervention strategies, to help slow the increasing T2D incidence in Malaysia.

Table 1: Demographic and clinical characteristics of the Malaysian sample (N=4077)

	Malays (N=1323)		Chinese (N=1344)		Indians (N=1410)	
	Control	T2D	Control	T2D	Control	T2D
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Gender						
Male	252 (34.85)	285 (47.5)***	153 (22.17)	343 (52.45)***	221 (31.48)	401 (56.64)***
Female	471 (65.15)	315 (52.5)	537 (77.83)	311 (47.55)	481 (68.52)	307 (43.36)
Age group, years						
Less than 50	424 (58.64)	200 (33.33)***	434 (62.90)	215 (32.87)***	486 (69.23)	293 (41.38)***
50-60	243 (33.61)	321 (53.50)	209 (30.29)	265 (40.52)	174 (24.79)	337 (47.60)
More than 60	56 (7.75)	79 (13.17)	47 (6.81)	174 (26.61)	42 (5.98)	78 (11.02)
BMI category, kg/m2						
Normal (<25)	266 (36.79)	153 (25.5)***	510 (73.91)	260 (39.76)***	271 (38.6)	241 (34.04)
Pre-obese (25-29.9)	313 (43.29)	268 (44.67)	152 (22.03)	267 (40.83)	275 (39.17)	292 (41.24)
Obese (>30)	144 (19.92)	179 (29.83)	28 (4.06)	127 (19.42)	156 (22.22)	175 (24.72)
Waist-to-Hip Ratio						
Low risk (<0.95 M, <0.80 F)	392 (54.22)	223 (37.17)***	407 (58.99)	232 (35.47)***	315 (44.87)	207 (29.24)***
Moderate risk (0.96-1 M, 0.81-0.85 F)	151 (20.89)	104 (17.33)	166 (24.06)	127 (19.42)	177 (25.21)	141 (19.92)
High risk (>1 M, >0.85 F)	180 (24.90)	273 (45.50)	117 (16.96)	295 (45.11)	210 (29.91)	360 (50.85)
Physical activity[a]						
Active	53 (15.01)	44 (10.50)	31 (7.73)	28 (8.28)	83 (15.29)	74 (17.41)
Inactive	300 (84.99)	375 (89.50)	370 (92.27)	310 (91.72)	460 (84.71)	351 (82.59)

Frequency of deep frying						
Less than once a week	76 (10.51)	80 (13.33)*	159 (23.04)	150 (22.94)	167 (23.79)	165 (23.31)
1-3x/week	245 (33.89)	231 (38.50)	205 (29.71)	212 (32.42)	194 (27.64)	215 (30.37)
4 or more times/week	402 (55.60)	289 (48.17)	326 (47.25)	292 (44.65)	341 (48.58)	328 (46.33)
Frequency of sauté						
Less than once a week	266 (36.79)	203 (33.83)	579 (83.91)	514 (78.59)*	402 (57.26)	424 (59.89)
1-3x/week	330 (45.64)	280 (46.67)	95 (13.77)	118 (18.04)	202 (28.77)	184 (25.99)
4 or more times/week	127 (17.57)	117 (19.50)	16 (2.32)	22 (3.36)	98 (13.96)	100 (14.12)
Frequency of coconut milk used						
Less than once a week	162 (22.41)	155 (25.83)	112 (16.23)	130 (19.88)	319 (45.44)	356 (50.28)
1-3x/week	320 (44.26)	237 (39.50)	235 (34.06)	206 (31.50)	282 (40.17)	247 (34.89)
4 or more times/week	241 (33.33)	208 (34.67)	343 (49.71)	318 (48.62)	101 (14.39)	105 (14.83)
Frequency of coffee consumption						
Less than once a week	265 (36.65)	216 (36.00)	191 (27.68)	165 (25.23)	184 (26.21)	181 (25.56)
1-3x/week	90 (12.45)	69 (11.50)	84 (12.17)	77 (11.77)	57 (8.12)	64 (9.04)
4 or more times/week	368 (50.90)	315 (52.50)	415 (60.14)	412 (63.00)	461 (65.67)	463 (65.40)

[a]N due to missing data: Malays=771; Chinese=739; Indians=968**.** Denotes statistically significant at *P<0.05; ** P<0.01; ***P<0.001

Table 2: Odds ratio and 95% confidence interval for primary analysis based on complete-case data (N=2478) and multiply imputed data (N=4077)

Group	Analysis	N	Odds ratio (95% CI)					Pseudo R^2 (%)
			Gender: Female (Ref: Male)	Age 50-60 (Ref: <50)	Age >60 (Ref: <50)	WHR	Physical inactivity (Ref: Physical activity)	
Malays	Complete-case	771	1.18 (0.81, 1.70); *P*=0.39	2.61 (1.88, 3.65); *P*<0.001	1.99 (1.07, 3.70); *P*=0.03	3.11 (2.39, 4.04); *P*<0.001	1.30 (0.79, 2.12); *P*=0.30	15.1
	Multiple Imputation	1323	1.14 (0.87, 1.49); *P*=0.34	2.25 (1.74, 2.90); *P*<0.001	1.98 (1.31, 2.99); *P*=0.001	3.19 (2.62, 3.87); *P*<0.001	1.27 (0.84, 1.93); *P*=0.25	15.0
Chinese	Complete-case	739	1.53 (0.98, 2.37); *P*=0.06	1.34 (0.92, 1.96); *P*=0.13	3.57 (2.08, 6.14); *P*<0.001	5.81 (4.19, 8.05); *P*<0.001	1.12 (0.59, 2.13); *P*=0.74	26.3
	Multiple Imputation	1344	1.19 (0.86, 1.66); *P*=0.29	1.70 (1.27, 2.28); *P*<0.001	3.93 (2.61, 5.93); *P*<0.001	6.37 (4.95, 8.19); *P*<0.001	1.00 (0.58, 1.73); *P*=0.99	30.3
Indians	Complete-case	968	1.10 (0.78, 1.56); *P*=0.58	2.32 (1.70, 3.16); *P*<0.001	2.13 (1.24, 3.64); *P*=0.006	3.60 (2.80, 4.63); *P*<0.001	1.07 (0.72, 1.60); *P*=0.73	18.5
	Multiple Imputation	1410	0.98 (0.74, 1.31); *P*=0.91	2.55 (1.97, 3.29); *P*<0.001	2.16 (1.39, 3.34); *P*=0.001	3.41 (2.79, 4.19); *P*<0.001	1.11 (0.74, 1.66); *P*=0.61	18.4
Combined[a]	Complete-case	2478	1.23 (0.99, 1.52); *P*=0.07	2.08 (1.72, 2.53); *P*<0.001	2.73 (1.98, 3.77); *P*<0.001	3.93 (3.36, 4.60); *P*<0.001	1.15 (0.87, 1.51); *P*=0.33	19.5
	Multiple Imputation	4077	1.08 (0.92, 1.28); *P*=0.34	2.18 (1.87, 2.54); *P*<0.001	2.75 (2.17, 3.50); *P*<0.001	4.01 (3.55, 4.53); *P*<0.001	1.13 (0.88, 1.45); *P*=0.32	20.5

[a] Combined model including ethnicity as a fixed effect.

Table 3: Area under the receiver operating characteristic curves (AUROC) with 95% confidence interval (CI) for each ancestral group, and comparisons between groups.

AUROC (95% CI)	**Malays**: 0.75 (0.72, 0.78)	**Chinese**: 0.83 (0.80, 0.86)	**Indians**: 0.78 (0.75, 0.81)	**Compared all 3 ancestry groups**
Malays: 0.75 (0.72, 0.78)		*P*<0.001	*P*=0.21	
Chinese: 0.83 (0.80, 0.86)	*P*<0.001		*P*=0.02	*P*=0.003
Indians: 0.78 (0.75, 0.81)	*P*=0.21	*P*=0.02		

Model included gender, age, waist-to-hip ratio (multiplied by 10) and physical activity

Figure 1 Forest plot of predictive model from primary complete-case analysis (N=2478)

Figure 2 Receiver operating characteristic (ROC) curves with 95% CI observed in each Malaysian ancestral group using primary complete-case analysis (N=2478). AUROC for combined group: 0.79 (0.77, 0.81)

References

Abdullah N, et al. (2015) Characterizing the genetic risk for Type 2 diabetes in a Malaysian multi-ethnic cohort. Diabetic medicine : a journal of the British Diabetic Association doi:10.1111/dme.12735

Abdullah N. AJ, Oldmeadow C., Scott R.J. & Holliday E.G (2014) The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings. International Journal of Endocrinology 2014:21

Alberti KG, Zimmet P, Shaw J (2007) International Diabetes Federation: a consensus on Type 2 diabetes prevention. Diabetic medicine : a journal of the British Diabetic Association 24(5):451-63 doi:10.1111/j.1464-5491.2007.02157.x

Cahill LE, et al. (2014) Fried-food consumption and risk of type 2 diabetes and coronary artery disease: a prospective study in 2 cohorts of US women and men. The American journal of clinical nutrition 100(2):667-675 doi:10.3945/ajcn.114.084129

Chambers JC, et al. (2015) Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol 3(7):526-34 doi:10.1016/S2213-8587(15)00127-8

Chan JC, et al. (2009) Diabetes in Asia: epidemiology, risk factors, and pathophysiology. JAMA : the journal of the American Medical Association 301(20):2129-40 doi:10.1001/jama.2009.726

Chen Z, et al. (2014) Joint effects of known type 2 diabetes susceptibility Loci in genome-wide association study of singapore chinese: the singapore chinese health study. PloS one 9(2):e87762 doi:10.1371/journal.pone.0087762

Cleves. MA (2002) From the help desk: Comparing areas under receiver operating characteristic curves from two or more probit or logit models. the Stata Journal 2(3):301-313

Population Distribution and Basic Demographic Characteristics 2010 (2011) Department of Statistics Malaysia,. Accessed 12 May 2014

Despres JP, Lemieux I (2006) Abdominal obesity and metabolic syndrome. Nature 444(7121):881-7 doi:10.1038/nature05488

Deurenberg P, Deurenberg-Yap M, Guricci S (2002) Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship. Obes Rev 3(3):141-6

Ewald B, McEvoy M, Attia J (2010) Pedometer counts superior to physical activity scale for identifying health markers in older adults. British journal of sports medicine 44(10):756-61 doi:10.1136/bjsm.2008.048827

Greenland S. Pearce N. (2014) Modeling strategies in epidemiology: II. Basic alternatives. Modeling Startegy in Epidemiology. p 1-20

Hu FB (2011) Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes care 34(6):1249-57 doi:10.2337/dc11-0442

International Diabetes Federation (2013) IDF Diabetes Atlas 6th Edition. vol 6, 6 edn. International Diabetes Federation, p 169

Jamal R. SZ, S.Z., Kamaruddin M.A., Jalal A.N, Ismail N., Kamil N.M, Abdullah N., Baharudin N., Hussin N.H, Othman H., Mahadi N. M., The Malaysian Cohort Group, (2014) Cohort profile: The Malaysian Cohort (TMC) project: a prospective study of non-communicable diseases in a multi-ethnic population. The International Journal of Epidemiology:9

Kaur P, et al. (2008) A comparison of anthropometric indices for predicting hypertension and type 2 diabetes in a male industrial population of Chennai, South India. Ethnicity & disease 18(1):31-6

Kelly P, Doherty A, Berry E, Hodges S, Batterham AM, Foster C (2011) Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. Int J Behav Nutr Phys Act 8:44 doi:10.1186/1479-5868-8-44

Knol MJ, VanderWeele TJ (2012) Recommendations for presenting analyses of effect modification and interaction. International journal of epidemiology 41(2):514-20 doi:10.1093/ije/dyr218

Landi F, Onder G, Bernabei R (2013) Sarcopenia and diabetes: two sides of the same coin. Journal of the American Medical Directors Association 14(8):540-1 doi:10.1016/j.jamda.2013.05.004

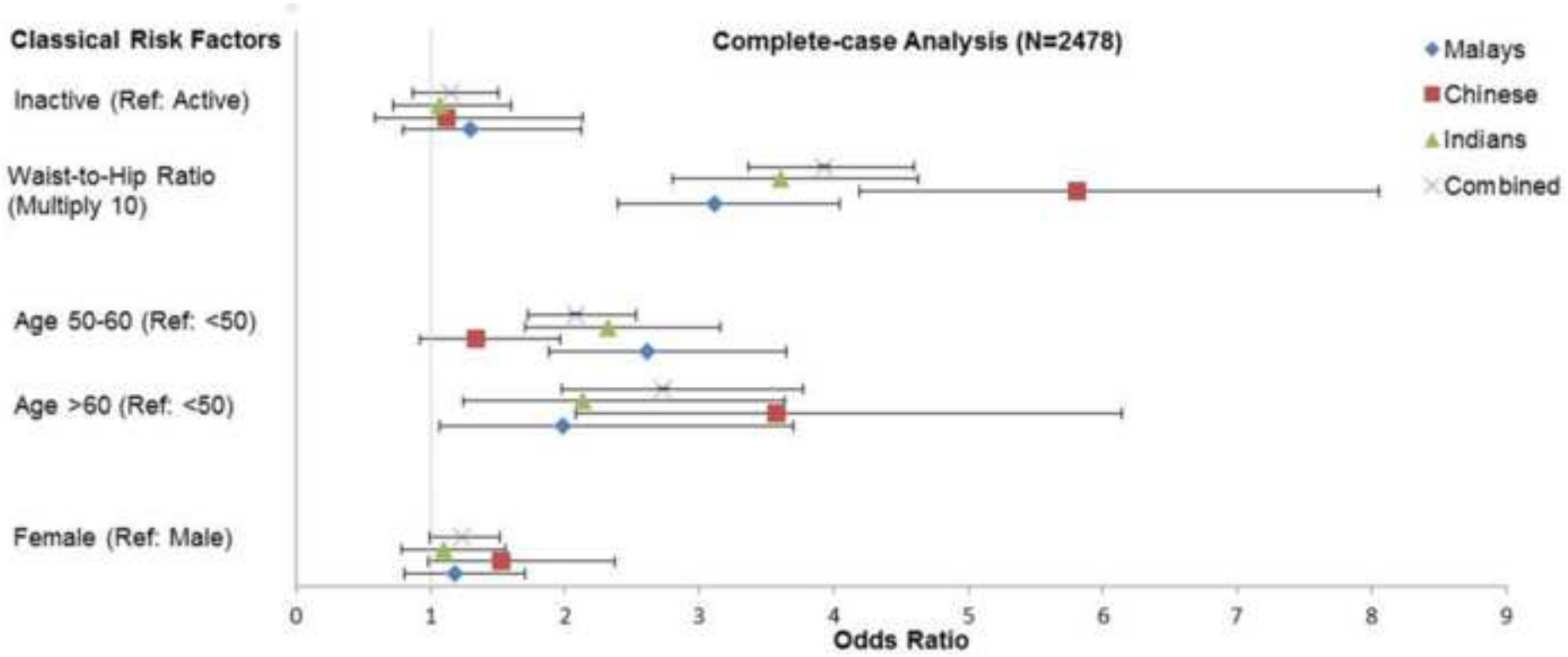
Lear SA, Humphries KH, Kohli S, Chockalingam A, Frohlich JJ, Birmingham CL (2007) Visceral adipose tissue accumulation differs according to ethnic background: results of the Multicultural Community Health Assessment Trial (M-CHAT). The American journal of clinical nutrition 86(2):353-9

Ling C, Groop L (2009) Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes 58(12):2718-25 doi:10.2337/db09-1003

Loopstra-Masters RC, Liese AD, Haffner SM, Wagenknecht LE, Hanley AJ (2011) Associations between the intake of caffeinated and decaffeinated coffee and measures of insulin sensitivity and beta cell function. Diabetologia 54(2):320-8 doi:10.1007/s00125-010-1957-8

Lopez-Garcia E, et al. (2005) Consumption of trans fatty acids is related to plasma biomarkers of inflammation and endothelial dysfunction. The Journal of nutrition 135(3):562-6
Lutsey PL, Pereira MA, Bertoni AG, Kandula NR, Jacobs DR, Jr. (2010) Interactions between race/ethnicity and anthropometry in risk of incident diabetes: the multi-ethnic study of atherosclerosis. American journal of epidemiology 172(2):197-204 doi:10.1093/aje/kwq100
Malik VS, Schulze MB, Hu FB (2006) Intake of sugar-sweetened beverages and weight gain: a systematic review. The American journal of clinical nutrition 84(2):274-88
Misra A, et al. (2011) Consensus dietary guidelines for healthy living and prevention of obesity, the metabolic syndrome, diabetes, and related disorders in Asian Indians. Diabetes Technol Ther 13(6):683-94 doi:10.1089/dia.2010.0198
Ng SW, Norton EC, Popkin BM (2009) Why have physical activity levels declined among Chinese adults? Findings from the 1991-2006 China Health and Nutrition Surveys. Soc Sci Med 68(7):1305-14 doi:10.1016/j.socscimed.2009.01.035
Odegaard AO, Koh WP, Yuan JM, Gross MD, Pereira MA (2012) Western-style fast food intake and cardiometabolic risk in an Eastern country. Circulation 126(2):182-8 doi:10.1161/CIRCULATIONAHA.111.084004
Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M (2008) A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 5:56 doi:10.1186/1479-5868-5-56
Qi Q, et al. (2014) Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. Bmj 348:g1610 doi:10.1136/bmj.g1610
Ronn T, et al. (2013) A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS genetics 9(6):e1003572 doi:10.1371/journal.pgen.1003572
Villegas R, et al. (2007) Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women. Archives of internal medicine 167(21):2310-6 doi:10.1001/archinte.167.21.2310
Wang Y, Mi J, Shan XY, Wang QJ, Ge KY (2007) Is China facing an obesity epidemic and the consequences? The trends in obesity and chronic disease in China. Int J Obes (Lond) 31(1):177-88 doi:10.1038/sj.ijo.0803354
Weeratunga P, Jayasinghe S, Perera Y, Jayasena G, Jayasinghe S (2014) Per capita sugar consumption and prevalence of diabetes mellitus--global and regional associations. BMC Public Health 14:186 doi:10.1186/1471-2458-14-186
World Health Organization (2008) Waist circumference and Waist-Hip Ratio. Report of a WHO Expert Consultation. Geneva, p 47
World Health Organization (2011) Global Recommendations on Physical Activity for Health 18-64 years old.
Wulan SN, Westerterp KR, Plasqui G (2010) Ethnic differences in body composition and the associated metabolic profile: a comparative study between Asians and Caucasians. Maturitas 65(4):315-9 doi:10.1016/j.maturitas.2009.12.012

S0378-5122(09)00467-8 [pii]
Xin Z, et al. (2012) Identifying obesity indicators which best correlate with type 2 diabetes in a Chinese population. BMC Public Health 12:732 doi:10.1186/1471-2458-12-732
Ye X, et al. (2014) Development of a new risk score for incident type 2 diabetes using updated diagnostic criteria in middle-aged and older chinese. PloS one 9(5):e97042 doi:10.1371/journal.pone.0097042



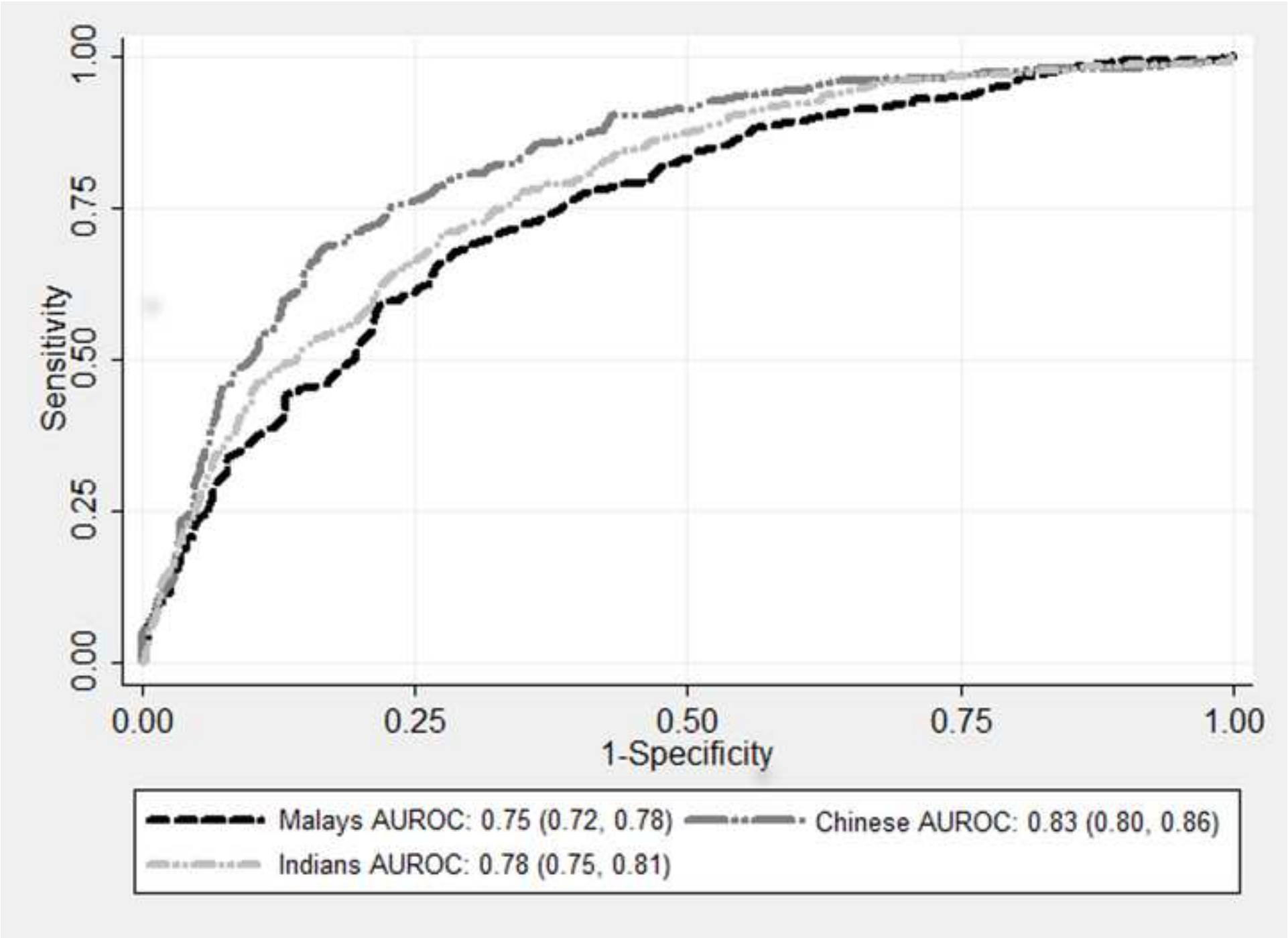


Table 1: Demographic and clinical characteristics of the Malaysian sample (N=4077)

	Malays (N=1323)		Chinese (N=1344)		Indians (N=1410)	
	Control	T2D	Control	T2D	Control	T2D
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Gender						
Male	252 (34.85)	285 (47.5)***	153 (22.17)	343 (52.45)***	221 (31.48)	401 (56.64)***
Female	471 (65.15)	315 (52.5)	537 (77.83)	311 (47.55)	481 (68.52)	307 (43.36)
Age group, years						
Less than 50	424 (58.64)	200 (33.33)***	434 (62.90)	215 (32.87)***	486 (69.23)	293 (41.38)***
50-60	243 (33.61)	321 (53.50)	209 (30.29)	265 (40.52)	174 (24.79)	337 (47.60)
More than 60	56 (7.75)	79 (13.17)	47 (6.81)	174 (26.61)	42 (5.98)	78 (11.02)
BMI category, kg/m2						
Normal (<25)	266 (36.79)	153 (25.5)***	510 (73.91)	260 (39.76)***	271 (38.6)	241 (34.04)
Pre-obese (25-29.9)	313 (43.29)	268 (44.67)	152 (22.03)	267 (40.83)	275 (39.17)	292 (41.24)
Obese (>30)	144 (19.92)	179 (29.83)	28 (4.06)	127 (19.42)	156 (22.22)	175 (24.72)
Waist-to-Hip Ratio						
Low risk (<0.95 M, <0.80 F)	392 (54.22)	223 (37.17)***	407 (58.99)	232 (35.47)***	315 (44.87)	207 (29.24)***
Moderate risk (0.96-1 M, 0.81-0.85 F)	151 (20.89)	104 (17.33)	166 (24.06)	127 (19.42)	177 (25.21)	141 (19.92)
High risk (>1 M, >0.85 F)	180 (24.90)	273 (45.50)	117 (16.96)	295 (45.11)	210 (29.91)	360 (50.85)
Physical activity[a]						
Active	53 (15.01)	44 (10.50)	31 (7.73)	28 (8.28)	83 (15.29)	74 (17.41)
Inactive	300 (84.99)	375 (89.50)	370 (92.27)	310 (91.72)	460 (84.71)	351 (82.59)

Frequency of deep frying						
Less than once a week	76 (10.51)	80 (13.33)*	159 (23.04)	150 (22.94)	167 (23.79)	165 (23.31)
1-3x/week	245 (33.89)	231 (38.50)	205 (29.71)	212 (32.42)	194 (27.64)	215 (30.37)
4 or more times/week	402 (55.60)	289 (48.17)	326 (47.25)	292 (44.65)	341 (48.58)	328 (46.33)
Frequency of sauté						
Less than once a week	266 (36.79)	203 (33.83)	579 (83.91)	514 (78.59)*	402 (57.26)	424 (59.89)
1-3x/week	330 (45.64)	280 (46.67)	95 (13.77)	118 (18.04)	202 (28.77)	184 (25.99)
4 or more times/week	127 (17.57)	117 (19.50)	16 (2.32)	22 (3.36)	98 (13.96)	100 (14.12)
Frequency of coconut milk used						
Less than once a week	162 (22.41)	155 (25.83)	112 (16.23)	130 (19.88)	319 (45.44)	356 (50.28)
1-3x/week	320 (44.26)	237 (39.50)	235 (34.06)	206 (31.50)	282 (40.17)	247 (34.89)
4 or more times/week	241 (33.33)	208 (34.67)	343 (49.71)	318 (48.62)	101 (14.39)	105 (14.83)
Frequency of coffee consumption						
Less than once a week	265 (36.65)	216 (36.00)	191 (27.68)	165 (25.23)	184 (26.21)	181 (25.56)
1-3x/week	90 (12.45)	69 (11.50)	84 (12.17)	77 (11.77)	57 (8.12)	64 (9.04)
4 or more times/week	368 (50.90)	315 (52.50)	415 (60.14)	412 (63.00)	461 (65.67)	463 (65.40)

[a]N due to missing data: Malays=771; Chinese=739; Indians=968**.** Denotes statistically significant at *P<0.05; ** P<0.01; ***P<0.001

Table 2: Odds ratio and 95% confidence interval for primary analysis based on complete-case data (N=2478) and multiply imputed data (N=4077)

Group	Analysis	N	Odds ratio (95% CI)					Pseudo R^2 (%)
			Gender: Female (Ref: Male)	Age 50-60 (Ref: <50)	Age >60 (Ref: <50)	WHR	Physical inactivity (Ref: Physical activity)	
Malays	Complete-case	771	1.18 (0.81, 1.70); *P*=0.39	2.61 (1.88, 3.65); *P*<0.001	1.99 (1.07, 3.70); *P*=0.03	3.11 (2.39, 4.04); *P*<0.001	1.30 (0.79, 2.12); *P*=0.30	15.1
	Multiple Imputation	1323	1.14 (0.87, 1.49); *P*=0.34	2.25 (1.74, 2.90); *P*<0.001	1.98 (1.31, 2.99); *P*=0.001	3.19 (2.62, 3.87); *P*<0.001	1.27 (0.84, 1.93); *P*=0.25	15.0
Chinese	Complete-case	739	1.53 (0.98, 2.37); *P*=0.06	1.34 (0.92, 1.96); *P*=0.13	3.57 (2.08, 6.14); *P*<0.001	5.81 (4.19, 8.05); *P*<0.001	1.12 (0.59, 2.13); *P*=0.74	26.3
	Multiple Imputation	1344	1.19 (0.86, 1.66); *P*=0.29	1.70 (1.27, 2.28); *P*<0.001	3.93 (2.61, 5.93); *P*<0.001	6.37 (4.95, 8.19); *P*<0.001	1.00 (0.58, 1.73); *P*=0.99	30.3
Indians	Complete-case	968	1.10 (0.78, 1.56); *P*=0.58	2.32 (1.70, 3.16); *P*<0.001	2.13 (1.24, 3.64); *P*=0.006	3.60 (2.80, 4.63); *P*<0.001	1.07 (0.72, 1.60); *P*=0.73	18.5
	Multiple Imputation	1410	0.98 (0.74, 1.31); *P*=0.91	2.55 (1.97, 3.29); *P*<0.001	2.16 (1.39, 3.34); *P*=0.001	3.41 (2.79, 4.19); *P*<0.001	1.11 (0.74, 1.66); *P*=0.61	18.4
Combined[a]	Complete-case	2478	1.23 (0.99, 1.52); *P*=0.07	2.08 (1.72, 2.53); *P*<0.001	2.73 (1.98, 3.77); *P*<0.001	3.93 (3.36, 4.60); *P*<0.001	1.15 (0.87, 1.51); *P*=0.33	19.5
	Multiple Imputation	4077	1.08 (0.92, 1.28); *P*=0.34	2.18 (1.87, 2.54); *P*<0.001	2.75 (2.17, 3.50); *P*<0.001	4.01 (3.55, 4.53); *P*<0.001	1.13 (0.88, 1.45); *P*=0.32	20.5

[a] Combined model including ethnicity as a fixed effect.

Table 3: Area under the receiver operating characteristic curves (AUROC) with 95% confidence interval (CI) for each ancestral group, and comparisons between groups.

AUROC (95% CI)	**Malays**: 0.75 (0.72, 0.78)	**Chinese**: 0.83 (0.80, 0.86)	**Indians**: 0.78 (0.75, 0.81)	**Compared all 3 ancestry groups**
Malays: 0.75 (0.72, 0.78)		*P*<0.001	*P*=0.21	*P*=0.003
Chinese: 0.83 (0.80, 0.86)	*P*<0.001		*P*=0.02	
Indians: 0.78 (0.75, 0.81)	*P*=0.21	*P*=0.02		

Model included gender, age, waist-to-hip ratio (multiplied by 10) and physical activity

Quantifying the Roles of Classical Risk Factors in Type 2 Diabetes using a Multi-ethnic Malaysian Cohort

Noraidatulakma Abdullah [1,2], Nor Azian Abdul Murad [2], John Attia [3,4], Christopher Oldmeadow [3,4], Mohd Arman Kamaruddin [2], Nazihah Abd Jalal [2], Norliza Ismail [2], Rahman Jamal [2], Rodney J. Scott [1,5], Elizabeth G. Holliday [3,4]

[1] School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia

[2] UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[3] Clinical Research Design, IT and Statistical Support (CReDITSS) Unit, Hunter Medical Research Institute, Newcastle, New South Wales, Australia

[4] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia

[5] Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales, Australia

Corresponding Authors:

Elizabeth G. Holliday

Mailing address:

CReDITSS, HMRI

C/- University Drive

Callaghan NSW 2308

Australia

E-mail: elizabeth.holliday@hmri.org.au

Telephone: +61(2)40420508

Facsimile: +61(2)40420039

Rahman Jamal

Mailing address:

UKM Medical Molecular Biology Institute (UMBI)

Jalan Yaacob Latif, Bandar Tun Razak, Cheras, 56000 Kuala Lumpur, Malaysia

E-mail: rahmanj@ppukm.ukm.edu.my

Telephone: +60(3)91459000

Facsimile: +60(3)91717185

Funding Sources

NA is supported by Ministry of Higher Education of Malaysia and Universiti Kebangsaan Malaysia. The Malaysian Cohort Study is funded by a top-down grant from the National Biotechnology Division, Ministry of Science, Technology and Innovation (MOSTI), Malaysia (ER-05-01-02-MEB001).

Conflict of Interest

We declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Acknowledgments

We thank all UKM Medical Molecular Biology Institute (UMBI) and The Malaysian Cohort staff members and research assistants. The voluntary participation of all the subjects is greatly appreciated.

4.4 Supplementary Data File for Publication 3

Data Supplement

Quantifying the Roles of Classical Risk Factors in Type 2 Diabetes using a Multi-ethnic Malaysian Cohort

International Journal of Public health

Noraidatulakma Abdullah [1,2], Nor Azian Abdul Murad [2], John Attia [3,4], Christopher Oldmeadow [3,4], Mohd Arman Kamaruddin [2], Nazihah Abd Jalal [2], Norliza Ismail [2], Rahman Jamal [2], Rodney J. Scott [1,5], Elizabeth G. Holliday [3,4]

[1] School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia

[2] UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[3] Clinical Research Design, IT and Statistical Support (CReDITSS) Unit, Hunter Medical Research Institute, Newcastle, New South Wales, Australia

[4] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, Faculty of Health, University of Newcastle, Newcastle, New South Wales, Australia

[5] Hunter Area Pathology Service, John Hunter Hospital, Newcastle, New South Wales, Australia

Corresponding Authors:

Elizabeth G. Holliday

Mailing address:

CReDITSS, HMRI

C/- University Drive

Callaghan NSW 2308

Australia

E-mail: elizabeth.holliday@hmri.org.au

Telephone: +61(2)40420508

Facsimile: +61(2)40420039

Rahman Jamal

Mailing address:

UKM Medical Molecular Biology Institute (UMBI)

Jalan Yaacob Latif, Bandar Tun Razak, Cheras, 56000 Kuala Lumpur, Malaysia

E-mail: rahmanj@ppukm.ukm.edu.my

Telephone: +60(3)91459000

Facsimile: +60(3)91717185

Online Resource 1: List of physical activities

List of moderate physical activities

At workplace:

1. Lifting moderate loads (5kg)
2. Lifting boxes
3. Mending
4. Hoeing weeds
5. Mowing
6. Gardening
7. Going up and downs stairs

At home:

1. Hoeing weeds
2. Gardening
3. Cleaning outside house
4. Raking
5. Mowing
6. Lifting moderate loads

Recreational:

1. Aerobic
2. Playing sport such as badminton, basketball, volleyball, bowling, table-tennis, golfing
3. Cycling
4. Dancing
5. Swimming
6. Brisk- walking
7. Yoga
8. Walking on treadmill
9. Qiqong
10. Tai-chi
11. Yoga
12. Heavy lifting gym

List of vigorous physical activities

At workplace:

1. Loading things into trucks
2. Lifting heavy things 7-18kg
3. Lifting heavy things upstairs
4. Using heavy tools (drilling, digging)
5. Digging trench

At home:

1. Moving and lifting furniture
2. Lifting things upstairs
3. Lifting heavy boxes
4. Hoeing weeds
5. Going ups and downs stairs

Recreational:

1. Aerobic
2. Cycling
3. Swimming
4. Playing sports such as badminton, football, volleyball, hockey, tennis, rugby
5. Jogging
6. Hiking
7. Martial arts, self-defence such as karate, judo, taekwondo

Online Resource 2: Demographic and clinical characteristics of 2478 complete-case data

	Malays (N=771)			Chinese (N=739)			Indians (N=968)			Combined (N=2478)		
	Control	**T2D**	**Total**	**Control**	**T2D**	**Total**	**Control**	**T2D**	**Total**	**Control**	**T2D**	**Total**
	N (%)	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**
Gender												
Male	129 (36.65)	205 (48.93)	334 (43.32)	99 (24.69)	169 (50.00)	268 (36.27)	181 (33.33)	244 (57.41)	425 (43.90)	409 (31.56)	618 (52.28)	1027 (41.44)
Female	223 (63.35)	214 (51.07)	437 (56.68)	302 (75.31)	169 (50.00)	471 (63.73)	362 (66.67)	181 (42.59)	543 (56.10)	887 (68.44)	564 (47.72)	1451 (58.56)
Age group, years												
Less than 50	231 (65.63)	159 (37.95)	390 (50.58)	238 (59.35)	119 (35.21)	357 (48.31)	379 (69.80)	180 (42.35)	559 (57.75)	848 (65.43)	458 (38.75)	1306 (52.70)
50-60	102 (28.98)	219 (52.27)	321 (41.63)	137 (34.16)	131 (38.76)	268 (36.27)	133 (24.49)	201 (47.29)	334 (34.50)	372 (28.70)	551 (46.62)	923 (37.25)
More than 60	19 (5.40)	41 (9.79)	60 (7.78)	26 (6.48)	88 (26.04)	114 (15.43)	31 (5.70)	44 (10.35)	75 (7.75)	76 (5.86)	173 (14.64)	249 (10.05)
BMI category, kg/m^2												
Normal (<25)	124 (35.23)	115 (27.45)	239 (31.00)	288 (71.82)	126 (37.28)	414 (56.02)	211 (38.86)	142 (33.41)	353 (36.47)	623 (48.07)	3838 (32.40)	1006 (40.60)
Pre-obese (25-29.9)	157 (44.60)	182 (43.44)	339 (43.97)	95 (23.69)	146 (43.20)	241 (32.61)	212 (39.04)	181 (42.59)	393 (40.60)	464 (35.80)	509 (43.06)	973 (39.27)
Obese (>30)	71 (20.17)	122 (29.12)	193 (25.03)	18 (4.49)	66 (19.53)	84 (11.37)	120 (22.10)	102 (24.00)	222 (22.93)	209 (16.13)	290 (24.53)	499 (20.14)
Waist-to-hip ratio												
Low risk (<0.95 M, <0.80 F)	189 (53.69)	163 (38.90)	352 (45.65)	234 (58.35)	112 (33.14)	346 (46.82)	246 (45.30)	121 (28.47)	367 (37.91)	669 (51.62)	396 (33.50)	1065 (42.98)
Moderate risk (0.96-1 M, 0.81-0.85 F)	83 (23.58)	80 (19.09)	163 (21.14)	102 (25.44)	64 (18.93)	166 (22.46)	143 (26.34)	83 (19.53)	226 (23.35)	328 (25.31)	227 (19.20)	555 (22.40)
High risk (>1 M, >0.85 F)	80 (22.73)	176 (42.00)	256 (33.20)	65 (16.21)	162 (47.93)	227 (30.72)	154 (28.36)	221 (52.00)	375 (38.74)	299 (23.07)	559 (47.29)	858 (34.62)

Waist circumference, cm												
Low risk (<93 M, <79 F)	196 (55.68)	175 (41.77)	371 (48.12)	287 (71.57)	139 (41.12)	426 (57.65)	244 (44.94)	156 (36.71)	400 (41.32)	727 (56.10)	470 (39.76)	1197 (48.31)
Moderate risk (94-101 M,80-79 F)	76 (21.59)	107 (25.54)	183 (23.74)	84 (20.95)	99 (29.29)	183 (24.76)	146 (26.89)	99 (23.29)	245 (25.31)	306 (23.61)	305 (25.80)	611 (24.66)
High risk (>102 M, >88 F)	80 (22.73)	137 (32.70)	217 (28.15)	30 (7.48)	100 (29.59)	130 (17.59)	153 (28.18)	170 (40.00)	323 (33.37)	263 (20.29)	407 (34.43)	670 (27.04)
Frequency of deep frying												
Less than once a week	36 (10.23)	58 (13.84)	94 (12.19)	94 (23.44)	75 (22.19)	169 (22.87)	129 (23.76)	101 (23.76)	230 (23.76)	259 (19.98)	234 (19.80)	493 (19.90)
1-3x/week	116 (32.95)	164 (39.14)	280 (36.32)	118 (29.43)	108 (31.95)	226 (30.58)	154 (28.36)	137 (32.24)	291 (30.06)	388 (29.94)	409 (34.60)	797 (32.16)
4 or more times/week	200 (56.82)	197 (47.02)	397 (51.49)	189 (47.13)	155 (45.86)	344 (46.55)	260 (47.88)	187 (44.00)	447 (46.18)	649 (50.08)	539 (45.60)	1188 (47.94)
Frequency of saute												
Less than once a week	135 (38.35)	139 (33.17)	274 (35.54)	331 (82.54)	267 (78.99)	598 (80.92)	305 (56.17)	256 (60.24)	561 (57.95)	771 (59.49)	662 (56.01)	1433 (57.83)
1-3x/week	151 (42.90)	197 (47.02)	348 (45.14)	60 (14.96)	61 (18.05)	121 (16.37)	162 (29.83)	115 (27.06)	277 (28.62)	373 (28.78)	373 (31.56)	746 (30.10)
4 or more times/week	66 (18.75)	83 (19.81)	149 (19.33)	10 (2.49)	10 (2.96)	20 (2.71)	76 (14.00)	54 (12.71)	130 (13.43)	152 (11.73)	147 (12.44)	299 (12.07)
Frequency of coconut milk used												
Less than once a week	85 (24.15)	114 (27.21)	199 (26.00)	72 (17.96)	82 (24.26)	154 (20.84)	244 (44.94)	199 (46.82)	443 (45.76)	401 (30.94)	395 (33.42)	796 (32.12)
1-3x/week	154 (43.75)	155 (36.99)	309 (40.08)	139 (34.66)	109 (32.25)	248 (33.56)	233 (42.91)	156 (36.71)	389 (40.19)	526 (40.59)	420 (35.53)	946 (38.18)
4 or more times/week	113 (32.10)	150 (35.80)	263 (34.11)	190 (47.38)	147 (43.49)	337 (45.6)	66 (12.15)	70 (16.47)	136 (14.05)	369 (28.47)	367 (31.05)	736 (29.7)
Frequency of coffee consumption												
Less than once a week	137 (38.92)	158 (37.71)	295 (38.26)	122 (30.42)	88 (26.04)	210 (28.42)	144 (26.52)	110 (25.88)	254 (26.24)	403 (31.10)	356 (30.12)	759 (30.63)
1-3x/week	46 (13.07)	55 (13.13)	101 (13.10)	50 (12.47)	37 (10.95)	87 (11.77)	46 (8.47)	45 (10.59)	91 (9.4)	142 (10.96)	137 (11.59)	279 (11.26)
4 or more times/week	169	206	375	229	213	442	353	270	623	751	689	1440

	(48.01)	(49.16)	(48.64)	(57.11)	(63.02)	(59.81)	(65.01)	(63.53)	(64.36)	(57.95)	(58.29)	(58.11)
Physical activity												
Active	53 (15.06)	44 (10.50)	97 (12.58)	31 (7.73)	28 (8.28)	59 (7.98)	83 (15.29)	74 (17.41)	157 (16.22)	167 (12.89)	146 (12.35)	313 (12.63)
Inactive	299 (84.94)	375 (89.50)	674 (87.42)	370 (92.27)	310 (91.72)	680 (92.02)	460 (84.71)	351 (82.59)	811 (83.78)	1129 (87.11)	1036 (87.65)	2165 (87.37)

Online Resource 3: Odds ratio and 95% confidence interval for secondary analysis (age, categorical waist-to-hip ratio and physical activity) based on complete-case and multiply imputed data

Group	Analysis	N	Odds ratio (95% CI)					Pseudo R^2 (%)
			Age 50-60 (Ref: <50)	Age >60 (Ref: <50)	WHR: Moderate risk (Ref: Low risk)	WHR: High risk (Ref: Low risk)	Physically inactive (Ref: Physically active)	
Malays	Complete-case	771	2.98 (2.17, 4.09); *P*<0.001*	2.91 (1.61, 5.25); *P*<0.001*	1.06 (0.72, 1.57); *P*=0.75	2.33 (1.63, 3.31); *P*<0.001*	1.12 (0.71, 1.77); *P*=0.61	8.0
	Multiple Imputation	1323	2.62 (2.05, 3.33); P<0.001*	2.66 (1.80, 3.93); *P*<0.001*	1.20 (0.88, 1.63); *P*=0.26	2.41 (1.86, 3.14); *P*<0.001*	0.98 (0.60, 1.61); *P*=0.94	
Chinese	Complete-case	739	1.59 (1.13, 2.25); *P*=0.008*	5.33 (3.20, 8.87); *P*<0.001*	1.24 (0.83, 1.85); *P*=0.29	4.40 (3.01, 6.43); *P*<0.001*	0.90 (0.50, 1.62); *P*=0.73	13.4
	Multiple Imputation	1344	2.19 (1.70, 2.83); *P*<0.001*	6.05 (4.17, 8.79); *P*<0.001*	1.24 (0.92, 1.67), *P*=0.15	3.60 (2.71, 4.77); *P*<0.001*	0.73 (0.45, 1.17); *P*=0.18	
Indians	Complete-case	968	2.91 (2.18, 3.89); *P*<0.001*	2.78 (1.68, 4.61); *P*<0.001*	1.10 (0.77, 1.58); *P*=0.60	2.60 (1.90, 3.55); *P*<0.001*	0.79 (0.55, 1.14); *P*=0.22	8.0
	Multiple Imputation	1410	3.02 (2.38, 3.84); *P*<0.001*	2.93 (1.95, 4.43); *P*<0.001*	1.15 (0.86, 1.54); *P*=0.35	2.39 (1.85, 3.09); *P*<0.001*	0.81 (0.58, 1.15); *P*=0.24	
Combined[a]	Complete-case	2478	2.49 (2.08, 2.98); *P*<0.001*	3.86 (2.84, 5.22); *P*<0.001*	1.13 (0.91, 1.41); *P*=0.27	2.91 (2.39, 3.55); *P*<0.001*	0.90 (0.70, 1.16); *P*=0.42	9.8
	Multiple Imputation	4077	2.63 (2.28, 3.02); *P*<0.001*	3.90 (3.12, 4.87); *P*<0.001*	1.20 (1.01, 1.43); *P*=0.04*	2.72 (2.34, 3.17); *P*<0.001*	0.83 (0.66, 1.06); *P*=0.14	

[a] Ethnic-adjusted model by fixed-effect. * denotes statistically significant at *P*<0.05.

Online Resource 4: Modification of the effect of ancestral groups on type 2 diabetes by waist-to-hip-ratio using complete-case data (N=2478)

		Ancestry groups						ORs (95% CI) for Chinese within strata of WHR	ORs (95% CI) for Indians within strata of WHR
		Malays		Chinese		Indians			
		N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)		
Waist-to-hip ratio	Lower risk	163/189	1.00	112/234	0.54 (0.39, 0.74); *P*<0.001*	121/246	0.61 (0.45, 0.84); *P*=0.002*	0.53 (0.38, 0.73); *P*<0.001*	0.61 (0.45, 0.84); *P*=0.002*
	Moderate	80/83	1.08 (0.74, 1.59); *P*=0.69	64/102	0.66 (0.45, 0.98); *P*=0.04*	83/143	0.67 (0.47, 0.96); *P*=0.03*	0.61 (0.39, 0.96); *P*=0.03*	0.62 (0.41, 0.95); *P*=0.03*
	High risk	176/80	2.39 (1.69, 3.38); *P*<0.001	162/65	2.33 (1.61, 3.37); *P*<0.001*	221/154	1.59 (1.17, 2.16); *P*=0.003*	1.01 (0.67, 1.51); *P*=0.97	0.66 (0.47, 0.94); *P*=0.02*
ORs (95% CI) for moderate risk WHR within strata of ancestry groups			1.06 (0.72, 1.57); *P*=0.75		1.24 (0.83, 1.85); *P*=0.29		1.10 (0.77, 1.58), *P*=0.60		
ORs (95% CI) for high risk WHR within strata of ancestry groups			2.33 (1.63, 3.31); *P*<0.001*		4.40 (3.01, 6.43); *P*<0.001*		2.60 (1.90, 3.55), *P*<0.001*		
Measure of effect modification on additive scale: RERI (95% CI) for moderate risk				1.04 (0.65, 1.66); *P*=0.86		0.98 (0.61, 1.57); *P*=0.93			
Measure of effect modification on additive scale: RERI (95% CI) for high risk				1.46 (0.57, 3.75); *P*=0.44		0.66 (0.29, 1.49); *P*=0.32			
Measure of effect modification on multiplicative scale: Ratio of ORs (95% CI) for moderate risk				1.13 (0.65, 1.96); *P*=0.67		1.02 (0.60, 1.72); *P*=0.95			
Measure of effect modification on multiplicative scale: Ratio of ORs (95% CI) for high risk				1.81 (1.08, 3.02); *P*=0.02*		1.09 (0.68, 1.73); *P*=0.72			

ORs are adjusted for age and physical activity. * denotes statistically significant at *P*<0.05.

Online Resource 5: Modification of the effect of ancestry groups on type 2 diabetes by age using complete-case data (N=2478)

		Ancestry groups						ORs (95% CI) for Chinese within strata of age	ORs (95% CI) for Indians within strata of age
		Malays		Chinese		Indians			
		N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)		
Age (years)	<50	159/231	1.00	119/238	0.78 (0.58, 1.07); *P*=0.12	180/379	0.64 (0.49, 0.84); *P*=0.002*	0.79 (0.58, 1.07); *P*=0.13	0.63 (0.48, 0.84); *P*=0.001*
	50-60	219/102	3.02 (2.20, 4.15); *P*<0.001*	131/137	1.30 (0.94, 1.80); *P*=0.11	201/133	1.86 (1.36, 2.53); *P*<0.001*	0.43 (0.31, 0.61); *P*<0.001*	0.63 (0.45, 0.88); *P*=0.006*
	>60	41/19	2.89 (1.59, 5.24); *P*<0.001*	88/26	4.32 (2.64, 7.08); *P*<0.001*	44/31	1.76 (1.05, 2.95); *P*=0.03*	1.57 (0.77, 3.20); *P*=0.22	0.63 (0.31, 1.31); *P*=0.22
ORs (95% CI) for aged 50-60 within strata of ancestry groups			2.98 (2.17, 4.09); *P*<0.001*		1.59 (1.13, 2.25); *P*=0.008*		2.91 (2.18, 3.89); *P*<0.001*		
ORs (95% CI) for aged >60 within strata of ancestry groups			2.91 (1.61, 5.25); *P*<0.001*		5.33 (3.20, 8.87); *P*<0.001*		2.78 (1.68, 4.61); *P*<0.001*		
Measure of effect modification on additive scale: RERI (95% CI) for aged 50-60				0.22 (0.09, 0.59); *P*=0.002*		0.47 (0.19, 1.17); *P*=0.11			
Measure of effect modification on additive scale: RERI (95% CI) for aged >60				5.06 (0.41, 62.5); *P*=0.21		0.47 (0.07, 3.00); *P*=0.42			
Measure of effect modification on multiplicative scale: Ratio of ORs (95% CI)for aged 50-60				0.55 (0.35, 0.87); *P*=0.01*		0.96 (0.62, 1.48); *P*=0.86			
Measure of effect modification on multiplicative scale: Ratio of ORs (95% CI)for aged >60				1.91 (0.88, 4.15); *P*=0.10		0.95 (0.43, 2.08); *P*=0.90			

ORs are adjusted for physical activity and waist-to-hip ratio. * denotes statistically significant at *P*<0.05.

Online Resource 6: Modification of the effect of ancestry groups on type 2 diabetes by physical activity using complete-case data (N=2478)

		Ancestry groups						ORs (95% CI) for Chinese within strata of physical activity	ORs (95% CI) for Indians within strata of physical activity
		Malays		Chinese		Indians			
		N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)		
Physical activity	Active	44/53	1.00	28/31	0.76 (0.38, 1.51); *P*=0.43	74/83	0.84 (0.49, 1.43); *P*=0.52	0.74 (0.36, 1.51); *P*=0.41	0.86 (0.49, 1.50); *P*=0.60
	Inactive	375/299	1.10 (0.70, 1.72); *P*=0.69	310/370	0.71 (0.45, 1.11); *P*=0.14	351/460	0.65 (0.42, 1.02); *P*=0.06	0.65 (0.52, 0.82); *P*<0.001*	0.59 (0.48, 0.74); *P*<0.001*
ORs (95% CI) for physical activity within strata of ancestry groups			1.12 (0.71, 1.77); *P*=0.61		0.90 (0.50, 1.62); *P*=0.73		0.79 (0.55, 1.14); *P*=0.22		
Measure of effect modification on additive scale: RERI (95% CI)				0.86 (0.44, 1.69); *P*=0.67		0.75 (0.40, 1.43); *P*=0.39			
Measure of effect modification on multiplicative scale: Ratio of ORs (95% CI)				0.85 (0.41, 1.77); *P*=0.67		0.71 (0.40, 1.26); *P*=0.24			

ORs are adjusted for age and waist-to-hip ratio. * denotes statistically significant at *P*<0.05.

Online Resource 7: The area under receiver operating curves (AUROC) and 95% confidence interval (CI) from secondary analysis (age, categorical waist-to-hip ratio and physical activity) based on complete-case and multiply imputed data

Group	Analysis	N	AUROC (95% CI)
Malays	Complete-case	771	0.75 (0.72, 0.78)
	Multiple Imputation	1323	0.75 (0.73, 0.78)
Chinese	Complete-case	739	0.83 (0.80, 0.86)
	Multiple Imputation	1344	0.85 (0.83, 0.87)
Indians	Complete-case	968	0.78 (0.75, 0.81)
	Multiple Imputation	1410	0.78 (0.75, 0.80)
Combined[a]	Complete-case	2478	0.79 (0.77, 0.81)
	Multiple Imputation	4077	0.79 (0.78, 0.81)

[a] Ethnic-adjusted model by fixed-effect.

Online Resource 8: Modification of the effect of ancestry groups on type 2 diabetes by gender using primary analysis (gender, age, continuous waist-to-hip ratio (multiply with 100) and physical activity)

		Ancestry groups						ORs (95% CI) for Chinese within strata of gender	ORs (95% CI) for Indians within strata of gender
		Malays		Chinese		Indians			
		N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)		
Gender	Male	205/129	1.00	169/99	0.86 (0.60, 1.25); *P*=0.43	244/181	0.58 (0.42, 0.80); *P*=0.001*	0.88 (0.61, 1.27); *P*=0.50	0.62 (0.44, 0.85); *P*=0.003*
	Female	214/223	1.41 (1.00, 1.98); *P*=0.05	169/302	0.96 (0.68, 1.36); *P*=0.82	181/362	0.67 (0.49, 0.93); *P*=0.02*	0.68 (0.50, 0.93); *P*=0.02*	0.46 (0.34, 0.63); *P*<0.001*
ORs (95% CI) for gender within strata of ancestry groups			1.18 (0.81, 1.70); *P*=0.39		1.53 (0.98, 2.37); *P*=0.06		1.10 (0.78, 1.56); *P*=0.58		
Measure of effect modification on additive scale: RERI (95% CI)				1.18 (0.81, 1.72); *P*=0.38		1.00 (0.81, 1.25); *P*=0.97			
Measure of interaction on multiplicative scale: Ratio of ORs (95% CI)				0.79 (0.49, 1.27); *P*=0.33		0.83 (0.53, 1.28); *P*=0.39			

ORs are adjusted for age, waist-to-hip ratio and physical activity. * denotes statistically significant at *P*<0.05.

Online Resource 9: Modification of the effect of ancestry groups on type 2 diabetes by age using primary analysis (gender, age, continuous waist-to-hip ratio (multiply with 10) and physical activity)

		Ancestry groups						ORs (95% CI) for Chinese within strata of age	ORs (95% CI) for Indians within strata of age
		Malays		Chinese		Indians			
		N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)		
Age (years)	<50	159/231	1.00	119/238	0.89 (0.64, 1.24); *P*= 0.49	180/379	0.55 (0.41, 0.74); *P*<0.001*	0.89 (0.64, 1.24); *P*= 0.48	0.55 (0.41, 0.75); *P*<0.001*
	50-60	219/102	2.62 (1.86, 3.68); *P*<0.001	131/137	1.25 (0.88, 1.78); *P*=0.20	201/133	1.26 (0.90, 1.75); *P*= 0.17	0.48 (0.33, .69); *P*<0.001*	0.47 (0.32, .67); *P*<0.001*
	>60	41/19	1.85 (0.99, 3.47); *P*=0.06	88/26	3.43 (2.05, 5.73); *P*<0.001*	44/31	1.15 (0.66, 2.00); *P*=0.61	1.80 (0.86, 3.78); *P*=0.12	0.64 (0.30, 1.36); *P*=0.12
ORs (95% CI) for aged 50-60 within strata of ancestry groups			2.61 (1.88, 3.65); *P*<0.001*		1.34 (0.92, 1.96); *P*=0.13		2.32 (1.70, 3.16); *P*<0.001*		
ORs (95% CI) for aged >60 within strata of ancestry groups			1.99 (1.07, 3.70); *P*=0.03*		3.57 (2.08, 6.14); *P*<0.001*		2.13 (1.24, 3.64); *P*=0.006*		
Measure of effect modification on additive scale: RERI (95% CI) for aged 50-60				0.29 (0.11, 0.73); *P*<0.009*		0.42 (0.18, 0.97); *P*=0.04*			
Measure of effect modification on additive scale: RERI (95% CI) for aged >60				5.47 (0.77, 38.96); *P*=0.09		0.74 (0.20, 2.78); *P*=0.65			
Measure of interaction on multiplicative scale: Ratio of ORs (95% CI)for aged 50-60				0.55 (0.35, 0.87); *P*=0.01*		1.91 (0.88, 4.15); *P*=0.10			
Measure of interaction on multiplicative scale: Ratio of ORs (95% CI)for aged >60				0.96 (0.62, 1.48); *P*=0.86		0.95 (0.43, 2.08); *P*=0.90			

ORs are adjusted for physical activity, gender and waist-to-hip ratio. * denotes statistically significant at *P*<0.05.

Online Resource 10: Modification of the effect of ancestry groups on type 2 diabetes by physical activity using primary analysis (gender, age, waist-to-hip ratio (multiply with 10) and physical activity)

		Ancestry groups						ORs (95% CI) for Chinese within strata of physical activity	ORs (95% CI) for Indians within strata of physical activity
		Malays		Chinese		Indians			
		N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)	N cases/controls	OR (95% CI)		
Physical activity	Active	44/53	1.00	28/31	0.87 (0.41, 1.82); *P*=0.70	74/83	0.64 (0.37, 1.13); *P*=0.13	0.84 (0.40, 1.77); *P*=0.65	0.75 (0.42, 1.34); *P*=0.70
	Inactive	375/299	1.33 (0.81, 2.16); *P*=0.26	310/370	0.98 (0.60, 1.59); *P*=0.93	351/460	0.67 (0.41, 1.09); *P*=0.10	0.74 (0.58, 0.96); *P*=0.02*	0.50 (0.39, 0.63); *P*<0.001*
ORs (95% CI) for physical activity within strata of ancestry groups			1.30 (0.79, 2.12); *P*=0.30		1.12 (0.59, 2.13); *P*=0.74		1.07 (0.72, 1.60); *P*=0.73		
Measure of effect modification on additive scale: RERI (95% CI)				0.81 (0.36, 1.80); *P*=0.61		0.74 (0.38, 1.47); *P*=0.39			
Measure of interaction on multiplicative scale: Ratio of ORs (95% CI)				0.85 (0.39, 1.86); *P*=0.68		0.79 (0.43, 1.48); *P*=0.46			

ORs are adjusted for age, gender and waist-to-hip ratio. * denotes statistically significant at *P*<0.05.

CHAPTER 5: PREDICTING TYPE 2 DIABETES USING GENETIC AND ENVIRONMENTAL RISK FACTORS IN A MULTI-ETHNIC MALAYSIAN COHORT

Abdullah, N., Abd Murad, NA., Mohamad Haniff, EA., Attia, J., Oldmeadow, C., Syafrudin, SE., Kamaruddin, MA., Ismail, N., Jalal, N., Ishak, M., Jamal, R., Holliday, E.

Public Health 149 (2017)31-38. April 2017. http://dx.doi.org/10.1016/j.puhe.2017.04.003 (PMID:28528225).

5.1 Statement of Co-authors

"As co-authors of the paper:

Abdullah, N., Abd Murad, NA., Mohamad Haniff, E A., Attia, J ., Oldmeadow, C ., Syafrudin, S E., Kamaruddin, MA., Ismail, N., Jalal, N., Ishak, M., Jamal, R., Holliday, E. Predicting Type 2 Diabetes using genetic and environmental risk factors in a multi-ethnic Malaysian Cohort. Public Health 149 (2017)31-38. April 2017. http://dx.doi.org/10.1016/j.puhe.2017.04.003 (PMID:28528225), we confirm that Noraidatulakma Abdullah contributed to this publication by performing the analysis, interpreting the result and writing the manuscript."

Co-Authors:

Dr Elizabeth Holliday
Signed
Date: 31 July 2017

Professor John Attia
Signed
Date: 1st August 2017

Dr Christopher Oldmeadow
Signed
Date: 1st August 2017

Professor Rodney Scott
Signed
Date: 2/8/2017

Professor Datuk A. Rahman A. Jamal
Signed
Date:

Dr Nor Azian Abdul Murad
Signed
Date:

Dr Ezanee Azlina Mohd Haniff
Signed
Date:

Saiful Effendi Syafruddin

Signed

Date:

Mohd Arman Kamaruddin

Signed

Date:

Nazihah Abd Jalal

Signed

Date:

Norliza Ismail

Signed

Date:

Muhiddin Ishak

Signed

Date:

Faculty Assistant Dean (Research Training)

Professor Robert Callister

Signed

Date: 3/08/2017

5.2 Summary of Publication 4

Introduction

Both genetic and non-genetic risk factors contribute to the pathogenesis of T2D. In prior studies in the Malaysian population, this research found that the contribution of known, common genetic factors to overall T2D risk was small (~2%), while environmental risk factors contributed substantially more (~20%). However, a substantial component of risk remains unexplained. Gene by environment interaction may contribute to the unknown component of variation, and also help to explain some of the "missing heritability". In this chapter, gene-environment interaction for T2D was assessed in the Malaysian population. Genetic risk factors comprised 62 known genetic risk variants, and a genetic risk score aggregating information across all variants. Environmental risk factors comprised those previously found to be associated with T2D in this Malaysian population.

PUBLIC HEALTH 149 (2017) 31–38

Available online at www.sciencedirect.com

Public Health

journal homepage: www.elsevier.com/puhe

Original Research

Predicting type 2 diabetes using genetic and environmental risk factors in a multi-ethnic Malaysian cohort

N. Abdullah [a,b], N.A. Abdul Murad [b], E.A. Mohd Haniff [b], S.E. Syafruddin [b], J. Attia [c,d], C. Oldmeadow [c,d], M.A. Kamaruddin [b], N. Abd Jalal [b], N. Ismail [b], M. Ishak [b], R. Jamal [b,**], R.J. Scott [a,e], E.G. Holliday [c,d,*]

[a] School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Newcastle, NSW, Australia

[b] UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[c] Clinical Research Design, IT and Statistical Support (CReDITSS) Unit, Hunter Medical Research Institute, Newcastle, NSW, Australia

[d] Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, Faculty of Health, University of Newcastle, Newcastle, NSW, Australia

[e] Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia

ARTICLE INFO

Article history:
Received 1 December 2016
Received in revised form
17 March 2017
Accepted 5 April 2017

Keywords:
Type 2 diabetes
Gene–environment interaction
Asian population
Epidemiology
Population studies

ABSTRACT

Objective: Malaysia has a high and rising prevalence of type 2 diabetes (T2D). While environmental (non-genetic) risk factors for the disease are well established, the role of genetic variations and gene–environment interactions remain understudied in this population. This study aimed to estimate the relative contributions of environmental and genetic risk factors to T2D in Malaysia and also to assess evidence for gene–environment interactions that may explain additional risk variation.

Study design: This was a case–control study including 1604 Malays, 1654 Chinese and 1728 Indians from the Malaysian Cohort Project.

Methods: The proportion of T2D risk variance explained by known genetic and environmental factors was assessed by fitting multivariable logistic regression models and evaluating McFadden's pseudo R^2 and the area under the receiver-operating characteristic curve (AUC). Models with and without the genetic risk score (GRS) were compared using the log likelihood ratio Chi-squared test and AUCs. Multiplicative interaction between genetic and environmental risk factors was assessed via logistic regression within and across ancestral groups. Interactions were assessed for the GRS and its 62 constituent variants.

Results: The models including environmental risk factors only had pseudo R^2 values of 16.5–28.3% and AUC of 0.75–0.83. Incorporating a genetic score aggregating 62 T2D-associated risk variants significantly increased the model fit (likelihood ratio *P*-value of 2.50×10^{-4}–4.83×10^{-12}) and increased the pseudo R^2 by about 1–2% and AUC by 1–3%. None of the

* *Corresponding author*. CReDITSS, HMRI, C/-University Drive, Callaghan, NSW 2308, Australia. Tel.: +61 (2) 40420508; fax: +61 (2) 40420039.

** *Corresponding author*. UKM Medical Molecular Biology Institute (UMBI), Jalan Yaacob Latif, Bandar Tun Razak, Cheras, 56000 Kuala Lumpur, Malaysia. Tel.: +60 (3) 91459000; fax: +60 (3) 91717185.

E-mail addresses: rahmanj@ppukm.ukm.edu.my (R. Jamal), elizabeth.holliday@hmri.org.au (E.G. Holliday).

http://dx.doi.org/10.1016/j.puhe.2017.04.003

gene–environment interactions reached significance after multiple testing adjustment, either for the GRS or individual variants. For individual variants, 33 out of 310 tested associations showed nominal statistical significance with $0.001 < P < 0.05$.

Conclusion: This study suggests that known genetic risk variants contribute a significant but small amount to overall T2D risk variation in Malaysian population groups. If gene–environment interactions involving common genetic variants exist, they are likely of small effect, requiring substantially larger samples for detection.

Introduction

Type 2 diabetes (T2D) is a complex polygenic disease influenced by both genetic and environmental risk factors. It has a high and rising prevalence, particularly in Asian countries. This increase seems to be largely attributable to environmental and lifestyle risk factors, resulting from substantial socio-economic growth and urbanisation.[1] Malaysia, a multi-ethnic country with a population of 28.3 million[2] has one of the highest comparative prevalences of T2D among Asian countries, with the prevalence continuing to rise.[3] T2D in this population has been relatively understudied compared to other Asian groups. The prevalence of T2D in Malaysia appears to differ among the three major Malaysian ancestral groups with Malaysian Indians having the highest prevalence (25–28%), followed by Malays (17–19%) and Chinese (9–14%).[4]

In addition to the contribution of lifestyle factors, T2D also has a substantial genetic component with heritability estimates in the order of 30–70%.[5,6] Although genome-wide association studies (GWAS) have identified hundreds of common variants associated with human diseases and traits, including T2D,[7] the most reported variants have small to moderate effects and individually account for only a small proportion of T2D heritability.[8] One important factor likely contributing to the 'missing heritability' is the disease's polygenic architecture, involving numerous genetic risk variants of individually small effect; many of which remain undetected in available samples.[9] Hence, despite heritability estimates of 30–70% for T2D, known variants appear to explain a minority of total genetic risk variation less than 10% in either European or Asian populations.[8,10,11] Alternatively, known lifestyle and environmental risk factors such as socio-demographic and measures of obesity account for a higher proportion of disease risk in populations.[12,13] For example, a study in a Dutch population found that lifestyle factors including smoking, alcohol consumption, physical activity and educational level explained 7.8% risk of T2D while adiposity accounted for 23.5% of T2D risk.[14] A cross-sectional conducted using the Boston Area Community Health III Survey also found that a high proportion of T2D risk was explained by environmental and lifestyle/behavioural factors (38.9% and 21.8% in black and Hispanic ancestry, respectively).[13] Nonetheless, a substantial component of T2D risk variance remains unexplained by known genetic variants or lifestyle/environmental factors.

In addition to the individual effects of genetic and environmental risk factors, gene–environment interactions may contribute an important component of T2D risk variance. In concert with lifestyle-related factors, interactions between particular genetic variants and these lifestyle factors may be a contributor to the increasing prevalence of T2D in the Malaysian and wider populations.[15]

This study aimed to assess the relative contributions of environmental factors, genetic variants and gene × environment interactions to T2D in Malaysia. Our study utilised data relating to lifestyle risk factors and genome-wide genetic variation in a large multi-ethnic Malaysian sample. In Malay, Chinese and Indian Malaysian samples, we first investigated the potential increase in predictive utility resulting from incorporating a genetic risk score (GRS) into a model containing environmental risk factors only. We then assessed evidence for gene–environment interaction for the GRS and each of its 62 constituent genetic variants, within and across the three ancestral groups.

Methods

Data sources and study samples

The study sample was selected from the Malaysian Cohort Project (MCP), a prospective population-based cohort including 106,527 volunteers aged between 35 and 70 years.[4] This case–control study included T2D cases and controls from the three major Malaysian ancestral groups: Malay, Chinese and Indian. Subjects were recruited between April 2006 and September 2012 from regions across Malaysia. For the current study, participants with fasting plasma glucose (FPG) exceeding 7.5 mmol/l (or 126 mg/dl) were classified as T2D with ancestry-matched control subjects having FPG < 5.5 mmol/l (or 99 mg/dl) without a previous diagnosis of diabetes.

A total of 4077 samples selected from the MCP were used in this analysis: 1323 Malays (600 cases and 723 controls), 1344 Chinese (654 cases and 690 controls) and 1410 Indians samples (708 cases and 702 controls). For selection, ethnicity was defined using the self-reported ethnicity of the subject and their family for three preceding generations. The slightly differing numbers of cases and controls resulted from previous application of quality control (QC) procedures to genetic data.[16] All relevant ethical approvals for the MCP were

approved by the Institutional Review and Ethics Board of Universiti Kebangsaan, Malaysia, in accordance with Declaration of Helsinki. All subjects gave their written and informed consent for participation in the study.

Genotyping and QC

Samples were genotyped at the UKM Medical Molecular Biology Institute, Kuala Lumpur, Malaysia, using the MetaboChip array (Illumina Inc, USA).[17] Genotype calling was performed using Illumina GenomeStudio Software with a default quality score (GenCall) thresholds of ≥0.3 and ≥0.25 for overall single-nucleotide polymorphisms (SNPs) and individual genotypes, respectively. Manual QC of genotype data was performed using PLINK,[18] resulting in exclusion of SNPs with rare minor allele frequency or evidence of potential genotyping errors, as described previously in detail.[16] Genetic ancestry was assessed by principal components analysis using reference data from the Singaporean Genome Variation Project (SGVP) and EIGENSTRAT software.[19] The SGVP was used due to high similarity between the Singaporean and Malaysian populations.

Candidate SNP selection

We previously selected SNPs showing genome-wide significant association ($P < 5 \times 10^{-8}$) with T2D using the online catalogue of published GWAS[7] and a comprehensive review of T2D genetic associations.[20] An SNP is a genetic variation occurring at a single DNA nucleotide. Identified SNPs were selected for testing in our Malaysian sample if they were present on the Metabochip array, and passed QC in at least two of the three Malaysian population groups. For T2D-associated loci containing multiple associated SNPs, we selected a single lead SNP from the largest study.[16] The final set of candidate SNPs were in approximate linkage equilibrium, with all pairwise $r^2 < 0.5$ based on linkage disequilibrium in HapMap Chinese/Japanese combined reference data.[21]

GRS construction

A genetic score was formed as a weighted sum of reference alleles for each candidate SNP, with weights specified as the log odds ratio (OR; beta coefficient) reported in the original publication. If multiple studies had reported genome-wide significant association of a SNP, we used the effect estimate reported by the largest study. Scoring was performed using PLINK.[22]

Selection of environmental risk factors

We selected known clinical, demographic and anthropometric T2D risk factors based on evidence from previous studies[6,23–25] and availability in the MCP study. We collectively refer to these risk factors simply as 'environmental' risk factors, to distinguish them from genetic factors. Environmental risk factors were measured either using self-report questionnaires or anthropometric measurements. Self-report questionnaires were used to measure age, gender, current smoking (yes/no), frequency of deep-fried food consumption, frequency of drinking coffee and physical activity. Smoking was assessed by asking the participant whether they currently smoked or used tobacco. Dietary variables were measured by asking participants how often they had consumed deep-fried foods and drank coffee in the preceding week. Questions had five response choices which were categorised into three groups: less than once a week, one to three times a week and more than four times a week.[26] Self-reported physical activity was classified using self-reported average weekly vigorous activity over the last 4 months, which was categorised as either active or inactive using a threshold of 150 min per week.[27] Anthropometric measures comprising body mass index (BMI), waist circumference (WC) and waist-to-hip ratio (WHR) were derived as the average of three measurements obtained using a seca or Harpenden stadiometer.

Statistical analyses

Multivariable logistic regression modelling was previously performed to investigate associations between environmental risk factors and T2D. An initial model was fitted including all identified risk factors. To produce the most parsimonious model, we removed variables showing no evidence of association in any ancestral group (at $P < 0.2$), providing the removal of variables produced no substantive changes in the model.[28] Variables remaining in the final model were used to assess predictive utility and gene–environment interaction (described below).

Using the final multivariate model, we first estimated the increment in variance explained resulting from adding the GRS to the model including environmental risk factors only. Models with and without the GRS were compared by assessing the difference in McFadden's pseudo R^2, conducting a likelihood ratio Chi-squared test (LR test) for the nested models and comparing the area under the receiver-operating characteristic curve (AUC) using De Long's test.[29,30] The AUC measures the predictive power and goodness of fit of logistic models. It represents the accuracy with which a model can differentiate between two outcome categories, and thus measures the potential diagnostic utility of the model. An ideal test has an AUC of 1, whereas a process of random guessing would produce an AUC of 0.5. Values of about 0.8 or greater are often considered clinically useful. Comparisons of these statistics were performed across all ancestral groups, adjusting for ancestry as a fixed effect, and also in each group individually.

We then assessed evidence for multiplicative interaction between each continuous environmental risk factor and the genetic score, adjusting for age and gender. Interaction analyses were also performed within each group individually and across all groups combined. For each environmental risk factor, we used a significance threshold of 0.05 for evaluating the GRS–environment interaction. We also assessed evidence for a gradient of GRS effects across categorical strata of environmental risk factors, and in a single model, we assessed the global significance of the interaction. These analyses were performed via logistic regression across the three groups combined, adjusting for ethnicity, age and gender. All reported results are from complete case analyses. Sensitivity analyses using multiply imputed data were performed but

provided very similar results. In the interests of space, these were not reported.

Secondary analyses

As a secondary analysis, we assessed multiplicative interaction between individual SNPs and environmental risk factors using logistic regression in all ancestral groups combined, included ethnicity as a fixed effect. The adjusted significance threshold for these analyses was based on Bonferroni correction for 62 SNPs and five environmental risk factors (age, gender, BMI, WHR, WC and physical activity, assuming high correlation between WHR and WC) (62 × 5). The pre-specified, adjusted significance threshold was thus $\alpha = 0.05/310 = 1.6 \times 10^{-4}$.

Based on the available sample size, for an interaction OR of 1.5, an assumed population prevalence of 17% for T2D[4] and significance threshold of 0.05, we had 23%, 36% and 50% power to detect gene–environment interaction for risk alleles with frequency 0.1, 0.2 and 0.5, respectively.[31] For a true OR of 1.1, power was low, ranging from 6% to 7% across allele frequencies (Fig. 1). For power of 80% to detect gene–environment interaction for risk alleles with frequencies 0.2, 0.3 and 0.4, the interaction ORs were 2.0, 1.9 and 1.8, respectively. All analyses were performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink) and STATA 11.2 (Stata Corporation, College Station, Texas).

Results

The initial model for environmental (clinical, demographic and anthropometric) risk factors included terms for age, gender, smoking, frequency of deep-fried food consumption, frequency of drinking coffee, physical activity, BMI, WC and WHR. After model reduction, the final model ($n = 2478$) included gender, age, WHR and physical activity (Table 1). Missing data were substantively due to the physical activity variable, resulting from a transition between two versions of physical activity questionnaires during the study. The combination of gender, age, WHR and physical inactivity explained about 15.1% disease risk in Malays, 26.3% in Chinese and 18.5% in Indians based on the pseudo R^2 (Table 1). In the combined sample, these four factors explained an estimated 19.5% of T2D risk. The AUC for the final model was highest in Chinese (AUC: 0.83, 95% confidence interval [CI]: 0.80–0.86), followed by Indians (AUC: 0.78, 95% CI: 0.75–0.81) and Malays (AUC: 0.75–95% CI: 0.72–0.78) (Table 1). In the combined group, the AUC was 0.79 (0.77–0.81) (Table 1).

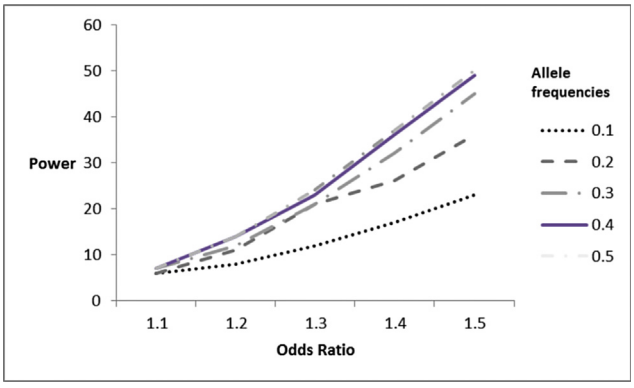


Fig. 1 – Power to detect gene–environment interaction by the range of odds ratio and allele frequencies.

A total of 62 T2D-associated SNPs were identified from the literature and available in our genetic data set, which has been previously described.[16] The addition of a GRS comprising these 62 SNPs to the environmental/lifestyle risk factor model increased the pseudo R^2 marginally by 1.4% in Malays (to 16.5%), 2.0% in Chinese (to 28.3%), 1.0% in Indians (to 19.6%) and 1.4% in combined samples (to 20.9%) (Table 1). The LR test comparing the two models indicated significant differences within the ancestral groups and in the combined group (Malays; $P < 1.53 \times 10^{-4}$, Chinese; $P < 7.80 \times 10^{-6}$, Indians; $P < 2.50 \times 10^{-4}$ and combined; $P < 4.83 \times 10^{-12}$). When comparing AUCs between the two models, only the AUC in Chinese (AUC: 0.83– 0.84; $P < 0.01$) and combined groups (AUC: 0.79– 0.80; $P < 0.001$) showed a significant increment as a result of adding the genetic score (Table 1). The AUC was slightly but non-significantly increased in Malays (increase by 0.02 unit: 0.77; $P = 0.06$) and Indians (increase by 0.01 unit: 0.79; $P = 0.15$) (Table 1).

For each lifestyle/environmental risk factor, Table 2 and Supplementary Table S1 (see Appendix A) show parameter estimates from models including main effects for the risk factor and the GRS (adjusted for age and gender) and a larger model that additionally incorporated an interaction term between the environmental factor and the GRS. The composite GRS demonstrated significant association with T2D in all models. All obesity parameters (BMI, WHR and WC) also showed highly significant association with T2D across the ancestral groups (Supplementary Table S1). Physical inactivity was only significantly associated with T2D in Malays (Supplementary Table S1). There was no evidence of multiplicative interaction between the GRS and any individual environmental risk factor either within or across the ancestral groups. The overall effect of the GRS across strata of environmental factors suggested a gradient of effects across strata of BMI (Table 3 and Fig. 2). The effect of the GRS was greatest in participants who were leaner (normal weight: OR = 1.49, 95% CI: 1.31–1.66), was smaller in overweight participants (OR = 1.38, 95% CI: 1.23–1.54) and was smallest in obese participants (OR = 1.21, 95% CI: 1.02–1.40), although the interaction term was not statistically significant ($P = 0.35$). There was no evidence of a gradient of GRS effects for other environmental risk factors.

Secondary analyses

There was no evidence of interaction between individual SNPs and any environmental/lifestyle risk factors after Bonferroni correction (i.e. at $P < 1.6 \times 10^{-4}$). A total 33 of 310 tested associations reached nominal significance ($P < 0.05$), with interaction P-values ranging from 0.001 to 0.05. The strongest associations were between BMI and rs10965250 within *CDKN2A* (OR: 1.61, 95% CI: 1.22–2.13, $P = 0.001$), rs4402960 within *IGF2BP2* (OR: 0.65, 95% CI: 0.50–0.85, $P = 0.001$) and

Table 1 – Comparing the risk explained and area under receiver-operating curves (AUCs) estimates for models with and without the genetic risk score within and across ancestral groups ($n = 2478$).

Groups	n	Pseudo R^2 (%)			AUC (95% CI)		
		Models without genetic score	Models with genetic score	LR test P-value	Models without genetic score	Models with genetic score	P-value for difference
Malays	771	15.1	16.5	1.53×10^{-4}*	0.75 (0.72–0.78)	0.77 (0.73–0.80)	0.06
Chinese	739	26.3	28.3	7.80×10^{-6}*	0.83 (0.80–0.86)	0.84 (0.82–0.87)	0.01*
Indians	968	18.5	19.6	2.50×10^{-4}*	0.78 (0.75–0.81)	0.79 (0.76–0.82)	0.15
Combined[a]	2478	19.5	20.9	4.83×10^{-12}*	0.79 (0.77–0.81)	0.80 (0.78–0.82)	0.001*

CI, confidence interval, LR, likelihood ratio Chi-squared test.
*Denotes statistical significance at $P < 0.05$.
[a] Ethnicity-adjusted model.

Table 2 – Comparing models with and without the interaction between genetic score and environmental risk factors in combined groups, coefficients are on the log-odds scale.

Group	Explaining variables	Model without interaction term			Model with interaction term		
		β	SE	P-value	β	SE	P-value
Combined[a]	BMI	0.535	0.037	<0.001*	0.540	0.038	<0.001*
	T2D 62 SNP-score	0.353	0.036	<0.001*	0.354	0.036	<0.001*
	BMI × T2D 62 SNP-score				−0.036	0.036	0.310
	WHR	1.176	0.053	<0.001*	1.179	0.053	<0.001*
	T2D 62 SNP-score	0.368	0.038	<0.001*	0.369	0.038	<0.001*
	WHR × T2D 62 SNP-score				−0.018	0.046	0.700
	WC	0.803	0.043	<0.001*	0.811	0.043	<0.001*
	T2D 62 SNP-score	0.372	0.037	<0.001*	0.372	0.037	<0.001*
	WC × T2D 62 SNP-score				−0.044	0.040	0.268
	Physical inactivity[b]	0.249	0.134	0.063	0.290	0.140	0.038*
	T2D 62 SNP-score	0.301	0.044	<0.001*	0.422	0.124	0.001*
	Physical inactivity[b] × T2D 62 SNP-score				−0.139	0.132	0.292

BMI, body mass index; WHR, waist-to-hip ratio; WC, waist circumference; T2D, type 2 diabetes; SNP, single-nucleotide polymorphism; SE, standard error.
*Denotes statistical significance at $P < 0.05$.
[a] Ethnicity-adjusted model.
[b] N = 4077 for BMI, WHR and WC; $n = 2478$ for physical inactivity due to missing data.

Table 3 – OR and 95% CI showing the effects of the genetic risk score within strata of environmental risk factors.

Characteristics	Category	By stratum	Interaction
		OR (95% CI)	Global P-value
Gender	Gender: male	1.31 (1.17–1.45)	0.39
	Gender: female	1.41 (1.28–1.54)	
Age in years	Age <50	1.36 (1.22–1.49)	0.23
	Age 50 to <60	1.31 (1.17–1.45)	
	Age ≥60	1.53 (1.19–1.87)	
WHR	WHR < (0.95 [M], 0.80 [F])	1.34 (1.20–1.48)	0.78
	WHR 0.95 to <0.99 (M), 0.80 to <0.84 (F)	1.32 (1.13–1.51)	
	WHR ≥ (0.99 [M], 0.84 [F])	1.35 (1.19–1.51)	
BMI	BMI < 25	1.49 (1.31–1.66)	0.35
	BMI 25 to <30	1.38 (1.23–1.54)	
	BMI ≥ 30	1.21 (1.02–1.40)	
WC[a]	WC < (94 [M], 80 [F])	1.38 (1.24–1.52)	0.93
	WC 94 to <102 (M), 80 to <88 (F)	1.3 (1.12–1.49)	
	WC ≥ (102 [M], 88 [F])	1.34 (1.16–1.52)	
Physical inactivity	Physically active (>150 min)	1.44 (1.08–1.80)	0.51
	Physically inactive (<150 min)	1.33 (1.20–1.45)	

OR, odds ratio; CI, confidence interval, M, male; F, female; WHR, waist-to-hip ratio; BMI, body mass index.
[a] Waist circumference, $n = 2478$ for all models.

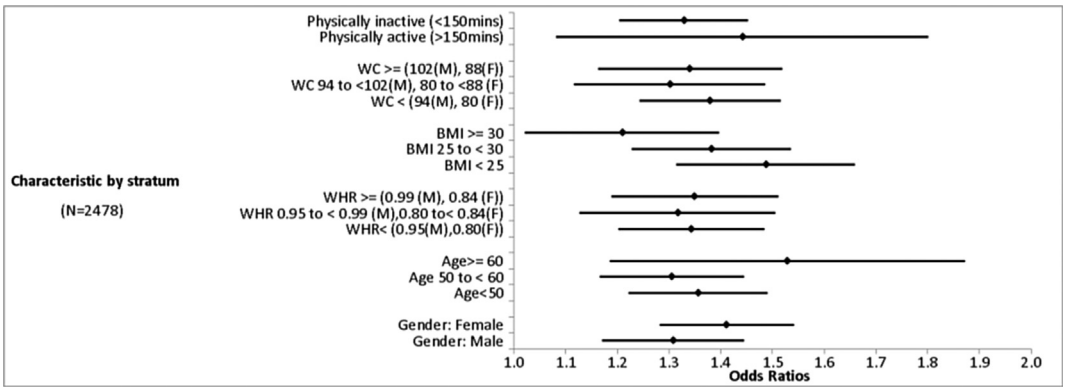


Fig. 2 – Type 2 diabetes increase in genetic risk score within strata. WC, waist circumference; BMI, body mass index; WHR, waist-to-hip ratio.

rs6808574 within *LPP* (OR: 0.42, 95% CI: 0.27–0.67, $P = 0.006$) (Supplementary Table S2a–Supplementary Table S2c).

Discussion

In a previous study, we found that four environmental risk factors accounted for about 20% of T2D risk across Malaysian groups. Here, we found that the addition of a genetic score into a model including environmental risk factors produced a small increase in variance explained both within and across the ancestry groups, with the increase ranging from 1% to 2%. Such a small increment produced by GRSs has also been observed in other studies of varied ancestry. These include a study of 3040 Han Chinese participants, which reported a 3% increment resulting from 10 risk variants, and a recent study that found a GRS comprising 62 SNPs explained 2.2% additional variance in white middle-aged adults ($n = 3471$), 1.5% in white young adults ($n = 1650$) and 1.6% in young black adults ($n = 820$).[32]

To our knowledge, this is the first large-scale study of T2D gene–environment interaction in Malaysia. Although we were unable to detect any significant interaction between a GRS and environmental risk factors, there was some indication of a gradient of genetic effects across BMI strata. Although we interpret this cautiously, it is interesting that a similar effect was found in the recent EPIC InterAct Case-Cohort study. This study included samples from nine European countries, and finding genetic effects for T2D were larger in participants who were leaner, both in terms of BMI (Hazard Ratio (HR): 1.6 in normal, 1.46 in overweight and 1.27 in obese) and WC (HR: 1.6 in low WC, 1.53 in medium WC and 1.29 in high WC).[33]

In addition, 33 tests reached nominal significance for interaction between individual SNPs and environmental risk factors, especially for BMI. Thus, true gene–environment interactions may exist but have eluded detection in our study.

There were several possible reasons underlying our inability to identify statistically significant interactions for either individual SNPs or the GRS. The first is a lack of statistical power. Detecting gene–environment interactions requires considerably larger samples than for detecting the corresponding main effects. Given the relatively small main effect of common T2D risk alleles, larger samples will likely be required to identify interactions. Our study had insufficient power to detect interaction effects on the order of those previously reported in gene–environment interaction studies (hazard ratios = 1.1–1.6).[34,35] Larger comprehensive studies in diverse populations will likely be necessary to provide sufficient power to detect gene–environment interactions involving individual variants.

Second, some measures of environmental exposure are difficult to quantify and standardise especially when based on self-report questionnaires, being prone to possible recall and reporting bias. Self-report questionnaires can introduce measurement error causing over- or under-reporting of exposures. Such measurement bias reduces statistical power. This is supported by a gene–environment interaction simulation study revealing that any moderate decreases in the measurement accuracy (correlation with true score of 0.4 vs 0.7) of the environmental risk factors can result in a 20-fold reduction in statistical power to detect interaction.[36] Such measurement error may also have reduced the estimated variance explained by the main effects of the studied environmental factors.

Another possibility is that genetic variants whose effect is modified by lifestyle factors could differ from the ones measured in this study. Therefore, approaches restricting attention to known loci may be non-exhaustive. Alternatively, there exist other important lifestyle risk factors that were not measured in this sample, including white rice and sweet beverage consumption that may have more important interactions with known T2D loci. Future studies may also reveal environmental influences on the transcriptome and metabolome that influence phenotypic variation.

Past studies interested in the diagnostic utility of genetic variation for T2D have generally not accounted for potential gene–environment interaction. Such studies have shown that known common risk variants for T2D offer only small improvements in risk prediction after considering established T2D risk factors. The utility of large-scale genotyping for disease prediction in a clinical setting has not been shown. Combining the information from common risk alleles with rare alleles of larger effect may, in the future, improve risk prediction of T2D.[37] It has been shown that patients with rare, monogenic version of diabetes have benefited from

personalised pharmacological treatment.[38] However, for the complex disease, T2D lifestyle factors currently appear to be better predictors of disease risk and targets for intervention. Interventions focused on reducing excess weight in all ancestral groups are urgently warranted in light of the current obesity epidemic.

In spite of numerous large-scale genetic studies, there is a substantial proportion of T2D risk unexplained. In order to explain the remaining heritability, important challenges remain. A major challenge is polygenic inheritance—many common risk variants will have such small effects that they will elude detection even in very large samples. For this reason, there will likely always be a component of the heritability that cannot be accounted for. More advanced, cost-effective genotyping and sequencing technologies in the future, combined with ongoing sample size enlargement, will also help to characterise other types of variation such as rare variation (including SNPs, deletions, duplications and inversions) and copy-number polymorphic duplication and assess their main effects, environmental interactions and contribution to variance explained.[9] Analytical developments for detecting complex patterns of association such as epistasis[39] may also help to shed light on the missing heritability. Studies of epigenetic inheritance such as DNA methylation and histone modification[40] might also help to close the gap of missing heritability.

In summary, we found only small effects of a GRS and no evidence for gene × environment interaction effects on T2D in a large multi-ethnic Malaysian sample. Interventions on known lifestyle risk factors are likely to offer the greatest utility for disease prevention and management for the foreseeable future.

Author statements

Acknowledgements

The authors thank all UKM Medical Molecular Biology Institute (UMBI) and the Malaysian Cohort staff members and research assistants. The voluntary participation of all the subjects is greatly appreciated.

Ethical approval

Ethical approvals for the Malaysian Cohort Project were approved by the Institutional Review and Ethics Board of Universiti Kebangsaan, Malaysia.

Funding

NA is supported by the Ministry of Higher Education of Malaysia and Universiti Kebangsaan Malaysia. The Malaysian Cohort Study is funded by a top-down grant from the National Biotechnology Division, Ministry of Science, Technology and Innovation (MOSTI), Malaysia (ER-05-01-02-MEB001).

Competing interests

None declared.

REFERENCES

1. Ramachandran A, Snehalatha C, Shetty AS, Nanditha A. Trends in prevalence of diabetes in Asian countries. *World J Diabetes* 2012;**3**:110–7.
2. *Population distribution and basic demographic characteristics 2010 [database on the Internet]*. Department of Statistics Malaysia; 2011 [cited 12 May 2014], Available from: https://www.dosm.gov.my/v1/index.php?r=column/ctheme&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09&bul_id=MDMxdHZjWTk1SjFzTzNkRXYzcVZjdz09.
3. International Diabetes Federation. *IDF diabetes atlas sixth edition poster update*. 2014.
4. Jamal R, Syed Zakaria SZ, Kamaruddin MA, Jalal AN, Ismail N, Kamil NM, et al., The Malaysian Cohort Group. Cohort profile: the Malaysian Cohort (TMC) project: a prospective study of non-communicable diseases in a multi-ethnic population. *Int J Epidemiol* 2014;**9**.
5. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance – a population-based twin study. *Diabetologia* 1999;**42**:139–45.
6. Carlsson S, Ahlbom A, Lichtenstein P, Andersson T. Shared genetic influence of BMI, physical activity and type 2 diabetes: a twin study. *Diabetologia* 2013;**56**:1031–5.
7. *A catalog of published genome-wide association studies [database on the Internet]*. National HUman Genome Research Institute; 2014 [cited April 2014]. Available from: www.genome.gov/gwastudies.
8. 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.
9. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012;**44**:483–9.
10. Chen Z, Pereira MA, Seielstad M, Koh WP, Tai ES, Teo YY, et al. Joint effects of known type 2 diabetes susceptibility loci in genome-wide association study of Singapore Chinese: the Singapore Chinese Health Study. *PLoS One* 2014;**9**:e87762.
11. Horikoshi M, Mgi R, van de Bunt M, Surakka I, Sarin AP, Mahajan A, et al. Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. *PLoS Genet* 2015;**11**:e1005230.
12. Hanley AJ, Festa A, D'Agostino Jr RB, Wagenknecht LE, Savage PJ, Tracy RP, et al. Metabolic and inflammation variable clusters and prediction of type 2 diabetes: factor analysis using directly measured insulin sensitivity. *Diabetes* 2004;**53**:1773–81.
13. Piccolo RS, Subramanian SV, Pearce N, Florez JC, McKinlay JB. Relative contributions of socioeconomic, local environmental, psychosocial, lifestyle/behavioral, biophysiological, and ancestral factors to racial/ethnic disparities in type 2 diabetes. *Diabetes Care* 2016;**39**:1208–17.
14. Abbasi A, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AM, van der AD, et al. Maternal and paternal transmission of type 2 diabetes: influence of diet, lifestyle and adiposity. *J Intern Med* 2011;**270**:388–96.
15. Eze IC, Imboden M, Kumar A, von Eckardstein A, Stolz D, Gerbase MW, et al. Air pollution and diabetes association: modification by type 2 diabetes genetic risk score. *Environ Int* 2016;**94**:263–71.
16. Abdullah N, Abdul Murad NA, Attia J, Oldmeadow C, Mohd Haniff EA, Syafruddin SE, et al. Characterizing the genetic risk for type 2 diabetes in a Malaysian multi-ethnic cohort. *Diabet Med* 2015;**32**:1377–84.

17. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. Correction: the metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* 2013;**9**.
18. Purcell S, Cherny SS, Sham PC. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;**19**:149–50.
19. Teo YY, Sim X, Ong RT, Tan AK, Chen J, Tantoso E, et al. Singapore genome variation project: a haplotype map of three Southeast Asian populations. *Genome Res* 2009;**19**:2154–62.
20. Abdullah N, Attia J, Oldmeadow C, Scott RJ, Holliday EG. The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings. *Int J Endocrinol* 2014;**2014**:21.
21. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;**24**:2938–9.
22. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;**460**:748–52.
23. van Dam RM, Feskens EJ. Coffee consumption and risk of type 2 diabetes mellitus. *Lancet* 2002;**360**:1477–8.
24. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino Sr RB. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007;**167**:1068–74.
25. Cahill LE, Pan A, Chiuve SE, Sun Q, Willett WC, Hu FB, et al. Fried-food consumption and risk of type 2 diabetes and coronary artery disease: a prospective study in 2 cohorts of US women and men. *Am J Clin Nutr* 2014;**100**:667–75.
26. Qi Q, Chu AY, Kang JH, Huang J, Rose LM, Jensen MK, et al. Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. *BMJ* 2014;**348**:g1610.
27. World Health Organization. *Global recommendations on physical activity for health 18–64 years old*. 2011.
28. Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. *Int J Epidemiol* 2016;**45**:565–75.
29. Cleves MA. From the help desk: comparing areas under receiver operating characteristic curves from two or more probit or logit models. *Stata J* 2002;**2**:301–13.
30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–45.
31. Gauderman W, Morrison J. *QUANTO 1.1: a computer program for power and sample size calculations for genetic-epidemiology studies*. 2006. http://hydra.usc.edu/gxe.
32. 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.
33. Langenberg C, Sharp SJ, Franks PW, Scott RA, Deloukas P, Forouhi NG, et al. Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. *PLoS Med* 2014;**11**:e1001647.
34. InterAct C, Romaguera D, Norat T, Wark PA, Vergnaud AC, Schulze MB, et al. Consumption of sweet beverages and type 2 diabetes incidence in European adults: results from EPIC-InterAct. *Diabetologia* 2013;**56**:1520–30.
35. InterAct C, Bendinelli B, Palli D, Masala G, Sharp SJ, Schulze MB, et al. Association between dietary meat consumption and incident type 2 diabetes: the EPIC-InterAct study. *Diabetologia* 2013;**56**:47–59.
36. Moffitt TE, Caspi A, Rutter M. Strategy for investigating interactions between measured genes and measured environments. *Arch Gen Psychiatry* 2005;**62**:473–81.
37. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 2008;**359**:2208–19.
38. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008;**299**:1335–44.
39. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. *Nat Rev Genet* 2014;**15**:722–33.
40. Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 2009;**460**:473–8.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.puhe.2017.04.003.

Data Supplement

Supplementary Table S1: Comparing models with and without the interaction between the genetic risk score and environmental risk factors across ancestral groups.

Groups	Explaining variables	Without interaction term			With interaction term		
		β	**SE**	***P*-value**	**β**	**SE**	***P*-value**
Malays (N=1323)	BMI[a]	0.4036	0.0618	<0.001*	0.4029	0.0616	<0.001*
	T2D 62 SNP-score	0.3435	0.0610	<0.001*	0.3476	0.0611	<0.001*
	BMI[a]*T2D 62 SNP-score				-0.0831	0.0585	0.1550
	WHR[b]	0.9592	0.0806	<0.001*	0.9592	0.0806	<0.001*
	T2D 62 SNP-score	0.3846	0.0645	<0.001*	0.3832	0.0646	<0.001*
	WHR[b]*T2D 62 SNP-score				0.0248	0.0773	0.7490
	WC[c]	0.6295	0.0678	<0.001*	0.6294	0.0679	<0.001*
	T2D 62 SNP-score	0.3658	0.0623	<0.001*	0.3677	0.0624	<0.001*
	WC[c]*T2D 62 SNP-score				-0.0718	0.0653	0.2710
	Physical Inactivity[d]	0.5112	0.2379	0.032*	0.5134	0.2396	0.032*
	T2D 62 SNP-score	0.2718	0.0777	<0.001*	0.2895	0.2319	0.2120

	Physical inactivity[d] *T2D 62 SNP-score				-0.0199	0.2460	0.9360
Chinese (N=1344)	BMI[a]	0.9637	0.0796	<0.001*	0.9640	0.0796	<0.001*
	T2D 62 SNP-score	0.3952	0.0685	<0.001*	0.3955	0.0685	<0.001*
	BMI[a]*T2D 62 SNP-score				-0.0089	0.0761	0.9060
	WHR[b]	1.6006	0.1112	<0.001*	1.6004	0.1113	<0.001*
	T2D 62 SNP-score	0.4054	0.0713	<0.001*	0.4051	0.0714	<0.001*
	WHR[b]*T2D 62 SNP-score				0.0059	0.0932	0.9490
	WC[c]	1.2701	0.0920	<0.001*	1.2697	0.0920	<0.001*
	T2D 62 SNP-score	0.3950	0.0709	<0.001*	0.3948	0.0709	<0.001*
	WC[c]*T2D 62 SNP-score				0.0461	0.0841	0.5840
	Physical inactivity[e]	0.2493	0.3044	0.4130	0.3054	0.3180	0.3370
	T2D 62 SNP-score	0.4069	0.0843	<0.001*	0.7119	0.3005	0.018*
	Physical inactivity[e] *T2D 62 SNP-score				-0.3329	0.3124	0.2870

Indians (N=1410)	BMI[a]	0.3228	0.0612	<0.001*	0.2933	0.0659	<0.001*
	T2D 62 SNP-score	0.3073	0.0604	<0.001*	0.3083	0.0605	<0.001*
	BMI[a]*T2D 62 SNP-score				0.0731	0.0614	0.2330
	WHR[b]	1.0674	0.0895	<0.001*	1.0913	0.0972	<0.001*
	T2D 62 SNP-score	0.3164	0.0635	<0.001*	0.3189	0.0636	<0.001*
	WHR[b]*T2D 62 SNP-score				-0.0518	0.0788	0.5110
	WC[c]	0.6259	0.0687	<0.001*	0.6318	0.0753	<0.001*
	T2D 62 SNP-score	0.3384	0.0622	<0.001*	0.3384	0.0622	<0.001*
	WC[c]*T2D 62 SNP-score				-0.0130	0.0676	0.8480
	Physical inactivity[f]	0.0426	0.1912	0.8240	0.0874	0.2112	0.6790
	T2D 62 SNP-score	0.2517	0.0707	<0.001*	0.3310	0.1710	0.0530
	Physical inactivity[f] *T2D 62 SNP-score				-0.0959	0.1878	0.6100

[a]Body Mass Index [b]Waist-to-hip ratio [c]Waist circumference [d]N=771 [e]N=739 [f]N=968. * denotes statistically significant at *P*<0.05

Supplementary Table S2a: Interaction between individual SNPs and the anthropometric measures body mass index (BMI) and waist-to-hip ratio (WHR)

Gene	Chr:BP	Lead SNPs	A1/A2	BMI 25-30 (Ref:<25) (N=3268)		BMI >30 (Ref:<25) (N=2376)		WHR 0.95-<0.99 (M), 0.80-<0.84 (F) (Ref:<0.95 (M), <0.80 (F)) (N=2642)		WHR >0.99 (M), >0.84 (F) (Ref:<0.95 (M), <0.80 (F)) (N=2301)	
				OR	P-value	OR	P-value	OR	P-value	OR	P-value
FAF1	1:50682573	rs17106184	A/G	1.46	[0.96,2.22]	1.12	[0.67,1.86]	0.98	[0.60,1.61]	0.76	[0.49,1.17]
NOTCH2	1:120319482	rs10923931	A/C	0.80	[0.57,1.10]	0.68*	[0.47,0.98]	0.93	[0.62,1.38]	0.94	[0.67,1.32]
LINC00538 - PROX1	1:212225879	rs340874	G/A	0.89	[0.68,1.16]	0.95	[0.68,1.32]	1.36	[0.98,1.89]	1.32*	[1.01,1.75]
GCKR	2:27594741	rs780094	A/G	1.02	[0.82,1.27]	1.02	[0.78,1.34]	0.86	[0.66,1.12]	1.00	[0.80,1.26]
THADA	2:43586327	rs7578597	G/A	0.78	[0.50,1.23]	0.44**	[0.26,0.73]	0.75	[0.43,1.31]	0.64	[0.40,1.01]
GRB14-COBLL1	2:165210095	rs13389219	A/G	0.79	[0.58,1.08]	0.90	[0.62,1.30]	0.91	[0.62,1.34]	0.75	[0.54,1.03]
LOC646736	2:226801989	rs7578326	G/A	1.02	[0.75,1.38]	0.87	[0.60,1.28]	0.89	[0.61,1.30]	1.02	[0.74,1.41]
PPARG	3:12264800	rs1801282	G/C	0.67	[0.42,1.07]	0.61	[0.35,1.05]	1.20	[0.68,2.13]	0.86	[0.53,1.39]
ADAMTS9-AS2	3:64686944	rs4607103	A/G	1.02	[0.77,1.35]	0.95	[0.66,1.36]	0.87	[0.62,1.23]	0.96	[0.72,1.29]
ADCY5	3:124548468	rs11708067	G/A	0.88	[0.59,1.32]	0.91	[0.57,1.46]	0.92	[0.56,1.53]	1.09	[0.72,1.65]
SLC2A2	3:172200215	rs11920090	A/T	0.82	[0.56,1.22]	0.75	[0.48,1.17]	1.65*	[1.03,2.64]	1.24	[0.83,1.86]
IGF2BP2	3:186994381	rs4402960	A/C	0.68**	[0.55,0.85]	0.65**	[0.50,0.85]	0.98	[0.75,1.29]	0.94	[0.75,1.18]
LPP	3:189223217	rs6808574	A/G	0.92	[0.61,1.41]	0.42***	[0.27,0.67]	0.78	[0.51,1.19]	0.69	[0.38,1.26]
WFS1	4:6353923	rs1801214	G/A	0.95	[0.71,1.27]	0.92	[0.66,1.30]	1.10	[0.77,1.57]	0.94	[0.70,1.28]
TMEM154	4:153739925	rs6813195	A/G	1.02	[0.82,1.26]	0.97	[0.75,1.26]	1.10	[0.84,1.43]	1.09	[0.87,1.36]
ARL15	5:53307177	rs702634	G/A	0.94	[0.72,1.23]	0.92	[0.67,1.26]	0.76	[0.55,1.06]	0.73*	[0.56,0.96]
ANKRD55-MAP3K1	5:55842508	rs459193	G/A	1.24	[0.84,1.82]	1.25	[0.81,1.95]	0.94	[0.60,1.45]	0.79	[0.54,1.16]
ZBED3-AS1	5:76460705	rs4457053	G/A	0.97	[0.72,1.33]	1.08	[0.74,1.57]	1.15	[0.79,1.68]	0.93	[0.67,1.29]
SSR1-RREB1	6:3672354	rs9505118	G/A	0.96	[0.78,1.19]	1.09	[0.84,1.41]	0.99	[0.76,1.29]	1.14	[0.91,1.43]
CDKAL1	6:20811931	rs6931514	G/A	1.13	[0.92,1.40]	1.11	[0.85,1.45]	1.07	[0.82,1.39]	1.27*	[1.01,1.59]

POU5F1-TCF19	6:31244432	rs3130501	A/G	0.93	[0.74,1.17]	0.86	[0.66,1.14]	0.87	[0.66,1.15]	1.05	[0.83,1.34]
VEGFA - C6orf223	6:43919740	rs9472138	A/G	1.29	[0.96,1.72]	1.39	[0.97,1.99]	0.93	[0.65,1.33]	0.88	[0.65,1.19]
DGKB - AGMO	7:15030834	rs2191349	C/A	1.06	[0.84,1.33]	0.84	[0.64,1.10]	0.92	[0.69,1.22]	0.87	[0.69,1.10]
JAZF1	7:28147081	rs864745	G/A	1.01	[0.79,1.30]	1.22	[0.90,1.65]	0.88	[0.65,1.20]	0.94	[0.73,1.22]
GCK - YKT6	7:44202193	rs4607517	A/G	0.89	[0.66,1.22]	1.29	[0.88,1.89]	1.00	[0.69,1.45]	0.93	[0.67,1.28]
ACHE	7:100328013	rs7636	A/G	1.34	[0.66,2.74]	0.84	[0.35,2.02]	0.89	[0.39,2.05]	0.85	[0.40,1.82]
FSCN3 - PAX4	7:127034139	rs10229583	A/G	0.92	[0.72,1.17]	0.85	[0.63,1.13]	0.92	[0.68,1.24]	0.98	[0.76,1.26]
KLF14 - MIR29A	7:130117394	rs972283	A/G	0.95	[0.76,1.19]	0.90	[0.69,1.17]	0.89	[0.68,1.16]	0.99	[0.79,1.25]
ANK1	8:41638405	rs516946	A/G	1.18	[0.86,1.60]	0.84	[0.58,1.23]	1.16	[0.79,1.69]	0.90	[0.65,1.24]
TP53INP1	8:96029687	rs896854	A/G	0.91	[0.72,1.14]	0.85	[0.65,1.11]	0.95	[0.72,1.26]	1.16	[0.92,1.47]
SLC30A8	8:118254206	rs3802177	A/G	1.29*	[1.04,1.60]	1.37*	[1.05,1.79]	0.99	[0.76,1.29]	0.96	[0.77,1.21]
GLIS3	9:4277466	rs7041847	A/G	1.06	[0.81,1.39]	1.10	[0.79,1.54]	1.00	[0.72,1.38]	0.85	[0.64,1.13]
CDKN2B-AS1 - DMRTA1	9:22123284	rs10965250	A/G	1.09	[0.87,1.36]	1.61***	[1.22,2.13]	0.96	[0.72,1.26]	1.08	[0.85,1.37]
TLE1-FAM75D5	9:83498768	rs2796441	G/A	1.03	[0.79,1.35]	1.08	[0.78,1.51]	0.87	[0.63,1.21]	0.98	[0.74,1.31]
CDC123 - MIR4480	10:12347900	rs11257655	A/G	1.02	[0.78,1.34]	1.21	[0.88,1.65]	1.15	[0.83,1.58]	1.18	[0.90,1.56]
VPS26A	10:70601480	rs1802295	A/G	0.68**	[0.51,0.90]	0.75	[0.53,1.07]	0.82	[0.57,1.17]	0.79	[0.58,1.06]
ZMIZ1	10:80612637	rs12571751	G/A	0.96	[0.78,1.19]	0.92	[0.71,1.20]	0.91	[0.70,1.19]	0.93	[0.74,1.17]
IDE - RPL11P4	10:94452862	rs6583826	G/A	0.90	[0.71,1.14]	0.77	[0.57,1.03]	0.75	[0.56,1.01]	0.90	[0.70,1.15]
HHEX - EXOC6	10:94337810	rs1111875	G/A	0.91	[0.73,1.14]	0.72*	[0.54,0.94]	1.12	[0.85,1.47]	1.07	[0.85,1.36]
ADRA2A - BTBD7P2	10:113032083	rs10885122	A/C	0.86	[0.64,1.15]	0.60**	[0.42,0.85]	0.74	[0.52,1.04]	0.84	[0.61,1.15]
TCF7L2	10:114748339	rs7903146	A/G	0.80	[0.60,1.07]	0.68*	[0.48,0.96]	1.05	[0.74,1.49]	0.98	[0.72,1.33]
KCNQ1; KCNQ1OT1	11:2648047	rs231362	A/G	0.73*	[0.55,0.97]	0.81	[0.57,1.15]	0.82	[0.58,1.17]	0.88	[0.65,1.19]
KCNJ11	11:17365206	rs5215	G/A	1.15	[0.92,1.43]	1.07	[0.81,1.40]	0.84	[0.64,1.11]	1.03	[0.82,1.30]
MADD	11:47292896	rs7944584	T/A	0.88	[0.61,1.27]	0.60*	[0.39,0.94]	1.07	[0.69,1.67]	0.78	[0.53,1.15]
FADS1	11:61328054	rs174550	A/G	1.15	[0.87,1.52]	1.30	[0.89,1.90]	1.43*	[1.01,2.03]	1.50**	[1.10,2.03]
ARAP1	11:72110746	rs1552224	C/A	0.83	[0.59,1.16]	0.91	[0.61,1.33]	0.81	[0.53,1.24]	0.80	[0.57,1.12]

FAT3 - MTNR1B	11:92313476	rs1387153	A/G	1.19	[0.97,1.48]	1.23	[0.95,1.60]	1.14	[0.87,1.48]	1.29*	[1.03,1.61]
KLHDC5	12:27856417	rs10842994	A/G	1.06	[0.81,1.40]	1.27	[0.92,1.76]	0.76	[0.55,1.06]	0.95	[0.72,1.26]
IGF1	12:101399699	rs35767	A/G	1.03	[0.82,1.29]	1.34*	[1.02,1.78]	0.88	[0.67,1.18]	0.93	[0.73,1.18]
RPL12P33 - HNF1A-AS1	12:122013881	rs7305618	G/A	1.03	[0.79,1.34]	0.85	[0.60,1.18]	1.16	[0.83,1.61]	0.98	[0.74,1.29]
MPHOSPH9	12:119887315	rs4275659	A/G	0.96	[0.77,1.20]	0.87	[0.66,1.13]	1.02	[0.78,1.34]	1.19	[0.95,1.50]
NDFIP2 - SPRY2	13:79615157	rs1359790	A/G	1.10	[0.84,1.44]	1.21	[0.87,1.67]	0.73	[0.52,1.02]	0.95	[0.72,1.25]
C2CD4A - C2CD4B	15:60201306	rs1436953	A/G	0.94	[0.76,1.17]	0.90	[0.69,1.16]	1.00	[0.76,1.31]	0.86	[0.69,1.08]
HMG20A-LINGO1	15:75619817	rs7177055	A/G	0.83	[0.67,1.03]	0.93	[0.72,1.21]	0.88	[0.68,1.14]	0.91	[0.72,1.14]
ZFAND6 - FAH	15:78219277	rs11634397	G/A	0.85	[0.58,1.23]	0.49**	[0.32,0.77]	1.18	[0.75,1.85]	0.91	[0.63,1.32]
PRC1; LOC100507118	15:89322341	rs8042680	C/A	1.13	[0.79,1.63]	0.97	[0.65,1.46]	1.26	[0.81,1.95]	0.83	[0.58,1.20]
FTO	16:52378028	rs9939609	A/T	0.76*	[0.59,0.96]	0.74*	[0.56,0.99]	1.08	[0.80,1.45]	0.94	[0.74,1.20]
CTRB2-CTRB1	16:73804746	rs7202877	C/A	1.37*	[1.02,1.84]	1.60**	[1.12,2.29]	1.08	[0.75,1.54]	1.18	[0.86,1.60]
SRR	17:2163008	rs391300	A/G	0.89	[0.68,1.16]	0.98	[0.70,1.37]	1.16	[0.84,1.60]	1.00	[0.75,1.32]
MC4R	18:56035730	rs12970134	A/G	1.12	[0.86,1.44]	0.91	[0.67,1.25]	1.14	[0.83,1.56]	1.28	[0.97,1.68]
PEPD	19:38584848	rs3786897	G/A	1.22	[0.98,1.52]	1.41*	[1.07,1.85]	1.00	[0.76,1.31]	1.14	[0.90,1.44]
HNF4A	20:42422681	rs4812829	A/G	1.03	[0.83,1.28]	1.24	[0.96,1.60]	1.02	[0.77,1.30]	0.95	[0.76,1.19]

Denotes statistically significant at *P<0.05; ** P<0.01; ***P<0.001

Supplementary Table S2b: Interaction of individual SNPs interaction with waist circumference (WC) and physical activity

Gene	Chr:BP	Lead SNPs	A1/A2	WC 94-<102 (M), 80-<88 (F) (Ref: <94 (M), <80 (F)) (N=2981)		WC >=102 (M), >=88 (F) (Ref: <94 (M), <80 (F)) (N=2087)		Physical activity (Ref:Active) (N=2478)	
				OR	P-value	OR	P-value	OR	P-value
FAF1	1:50682573	rs17106184	A/G	1.32	[0.84,2.08]	0.77	[0.48,1.23]	1.00	[0.49, 2.06]
NOTCH2	1:120319482	rs10923931	A/C	0.92	[0.64,1.32]	0.74	[0.53,1.04]	1.18	[0.71,1.96]
LINC00538 - PROX1	1:212225879	rs340874	G/A	1.29	[0.96,1.74]	1.06	[0.79,1.43]	1.01	[0.63,1.63]
GCKR	2:27594741	rs780094	A/G	1.15	[0.90,1.47]	1.04	[0.82,1.33]	1.20	[0.83,1.75]
THADA	2:43586327	rs7578597	G/A	0.62	[0.37,1.04]	0.49**	[0.31,0.78]	0.60	[0.33,1.07]
GRB14-COBLL1	2:165210095	rs13389219	A/G	0.93	[0.65,1.34]	0.80	[0.58,1.12]	1.09	[0.64,1.84]
LOC646736	2:226801989	rs7578326	G/A	0.77	[0.55,1.09]	0.84	[0.60,1.17]	1.13	[0.67,1.89]
PPARG	3:12264800	rs1801282	G/C	0.73	[0.43,1.23]	0.67	[0.41,1.11]	0.97	[0.47,2.01]
ADAMTS9-AS2	3:64686944	rs4607103	A/G	0.88	[0.64,1.20]	0.94	[0.68,1.29]	0.57	[0.32,1.02]
ADCY5	3:124548468	rs11708067	G/A	1.03	[0.66,1.60]	1.10	[0.72,1.67]	0.80	[0.44,1.46]
SLC2A2	3:172200215	rs11920090	A/T	1.17	[0.76,1.79]	0.82	[0.55,1.23]	0.72	[0.40,1.30]
IGF2BP2	3:186994381	rs4402960	A/C	0.82	[0.65,1.05]	0.83	[0.66,1.05]	1.22	[0.84,1.76]
LPP	3:189223217	rs6808574	A/G	0.82	[0.57,1.18]	1.04	[0.53,2.03]	0.95	[0.67,1.36]
WFS1	4:6353923	rs1801214	G/A	1.01	[0.74,1.38]	0.89	[0.65,1.22]	0.59*	[0.37,0.94]
TMEM154	4:153739925	rs6813195	A/G	1.12	[0.88,1.42]	1.06	[0.84,1.34]	1.18	[0.82,1.70]
ARL15	5:53307177	rs702634	G/A	0.84	[0.63,1.13]	0.86	[0.65,1.14]	1.01	[0.66,1.54]
ANKRD55-MAP3K1	5:55842508	rs459193	G/A	0.63*	[0.41,0.95]	1.05	[0.71,1.55]	0.80	[0.41,1.55]
ZBED3-AS1	5:76460705	rs4457053	G/A	0.79	[0.56,1.11]	1.09	[0.78,1.52]	1.11	[0.66,1.86]
SSR1-RREB1	6:3672354	rs9505118	G/A	1.11	[0.87,1.41]	1.12	[0.89,1.41]	0.91	[0.63,1.33]
CDKAL1	6:20811931	rs6931514	G/A	1.15	[0.90,1.47]	1.19	[0.94,1.51]	1.01	[0.71,1.43]
POU5F1-TCF19	6:31244432	rs3130501	A/G	0.96	[0.74,1.24]	0.88	[0.69,1.13]	1.09	[0.75,1.60]
VEGFA - C6orf223	6:43919740	rs9472138	A/G	1.08	[0.78,1.50]	1.05	[0.77,1.44]	1.33	[0.80,2.21]
DGKB - AGMO	7:15030834	rs2191349	C/A	0.89	[0.69,1.15]	0.94	[0.73,1.20]	0.93	[0.63,1.37]

JAZF1	7:28147081	rs864745	G/A	0.96	[0.73,1.27]	1.04	[0.79,1.36]	1.08	[0.70,1.65]
GCK - YKT6	7:44202193	rs4607517	A/G	0.88	[0.63,1.24]	1.18	[0.83,1.68]	0.74	[0.41,1.32]
ACHE	7:100328013	rs7636	A/G	1.05	[0.50,2.19]	1.22	[0.51,2.89]	1.07	[0.41,2.80]
FSCN3 - PAX4	7:127034139	rs10229583	A/G	0.79	[0.60,1.03]	0.80	[0.62,1.04]	0.83	[0.57,1.21]
KLF14 - MIR29A	7:130117394	rs972283	A/G	0.96	[0.75,1.23]	0.97	[0.77,1.23]	0.86	[0.59,1.24]
ANK1	8:41638405	rs516946	A/G	0.98	[0.69,1.39]	0.97	[0.69,1.36]	0.73	[0.41,1.28]
TP53INP1	8:96029687	rs896854	A/G	0.88	[0.68,1.13]	1.02	[0.80,1.31]	0.74	[0.51,1.08]
SLC30A8	8:118254206	rs3802177	A/G	0.99	[0.78,1.26]	1.24	[0.98,1.58]	1.01	[0.69,1.48]
GLIS3	9:4277466	rs7041847	A/G	0.91	[0.67,1.23]	0.92	[0.68,1.25]	0.92	[0.54,1.58]
CDKN2B-AS1 - DMRTA1	9:22123284	rs10965250	A/G	1.28	[1.00,1.65]	1.34*	[1.05,1.72]	1.44	[0.96,2.18]
TLE1-FAM75D5	9:83498768	rs2796441	G/A	1.07	[0.79,1.45]	1.03	[0.76,1.40]	0.81	[0.49,1.35]
CDC123 - MIR4480	10:12347900	rs11257655	A/G	1.01	[0.75,1.36]	1.13	[0.85,1.49]	1.44	[0.94,2.19]
VPS26A	10:70601480	rs1802295	A/G	0.63**	[0.46,0.88]	0.81	[0.59,1.11]	0.90	[0.57,1.44]
ZMIZ1	10:80612637	rs12571751	G/A	0.98	[0.77,1.25]	0.99	[0.78,1.26]	1.08	[0.76,1.52]
IDE - RPL11P4	10:94452862	rs6583826	G/A	0.89	[0.68,1.17]	0.79	[0.61,1.03]	1.44	[0.95,2.17]
HHEX - EXOC6	10:94337810	rs1111875	G/A	1.06	[0.82,1.37]	0.92	[0.72,1.17]	1.25	[0.86,1.83]
ADRA2A - BTBD7P2	10:113032083	rs10885122	A/C	0.70*	[0.51,0.97]	0.64**	[0.46,0.87]	0.87	[0.56,1.36]
TCF7L2	10:114748339	rs7903146	A/G	0.69*	[0.49,0.96]	0.88	[0.65,1.20]	0.67	[0.42,1.06]
KCNQ1; KCNQ1OT1	11:2648047	rs231362	A/G	0.78	[0.57,1.08]	0.80	[0.59,1.09]	0.87	[0.54,1.38]
KCNJ11	11:17365206	rs5215	G/A	1.01	[0.79,1.29]	1.13	[0.88,1.43]	1.07	[0.72,1.60]
MADD	11:47292896	rs7944584	T/A	0.90	[0.60,1.36]	0.70	[0.48,1.04]	0.87	[0.48,1.58]
FADS1	11:61328054	rs174550	A/G	0.98	[0.71,1.34]	1.19	[0.85,1.67]	1.24	[0.72,2.14]
ARAP1	11:72110746	rs1552224	C/A	0.99	[0.68,1.43]	0.78	[0.55,1.12]	0.90	[0.54,1.49]
FAT3 - MTNR1B	11:92313476	rs1387153	A/G	1.31*	[1.03,1.66]	1.19	[0.95,1.51]	0.99	[0.68,1.42]
KLHDC5	12:27856417	rs10842994	A/G	0.95	[0.71,1.29]	1.26	[0.93,1.70]	0.82	[0.52,1.30]
IGF1	12:101399699	rs35767	A/G	0.91	[0.70,1.17]	1.00	[0.78,1.28]	1.27	[0.85,1.89]
RPL12P33 - HNF1A-AS1	12:122013881	rs7305618	G/A	1.07	[0.80,1.43]	1.05	[0.78,1.41]	1.11	[0.66,1.86]

MPHOSPH9	12:119887315	rs4275659	A/G	1.08	[0.84,1.37]	1.10	[0.86,1.40]	1.08	[0.73,1.60]
NDFIP2 - SPRY2	13:79615157	rs1359790	A/G	1.13	[0.84,1.52]	1.13	[0.85,1.51]	0.94	[0.61, 1.46]
C2CD4A - C2CD4B	15:60201306	rs1436953	A/G	1.02	[0.80,1.30]	0.94	[0.74,1.18]	1.04	[0.72,1.50]
HMG20A-LINGO1	15:75619817	rs7177055	A/G	0.88	[0.69,1.12]	0.86	[0.68,1.09]	0.68*	[0.47,0.97]
ZFAND6 - FAH	15:78219277	rs11634397	G/A	0.80	[0.53,1.22]	0.55**	[0.37,0.81]	0.73	[0.38,1.40]
PRC1; LOC100507118	15:89322341	rs8042680	C/A	0.80	[0.54,1.19]	0.74	[0.51,1.07]	0.93	[0.56,1.55]
FTO	16:52378028	rs9939609	A/T	0.91	[0.70,1.18]	0.81	[0.62,1.04]	0.95	[0.64,1.39]
CTRB2-CTRB1	16:73804746	rs7202877	C/A	1.12	[0.81,1.55]	1.53*	[1.10,2.12]	1.10	[0.67,1.81]
SRR	17:2163008	rs391300	A/G	1.03	[0.77,1.37]	0.88	[0.65,1.19]	1.25	[0.76,2.05]
MC4R	18:56035730	rs12970134	A/G	0.85	[0.64,1.14]	1.11	[0.84,1.47]	1.39	[0.92,2.12]
PEPD	19:38584848	rs3786897	G/A	1.26	[0.98,1.62]	1.30*	[1.02,1.66]	0.79	[0.55,1.13]
HNF4A	20:42422681	rs4812829	A/G	1.02	[0.80,1.30]	1.23	[0.97,1.56]	0.94	[0.66,1.35]

Denotes statistically significant at *P<0.05; ** P<0.01; ***P<0.001

Supplementary Table S2c: Interaction of individual SNPs with gender and age

Gene	Chr:BP	Lead SNPs	A1/A2	Gender Female (Ref: Male) (N=4077)		Age 50-60 (Ref:<50) (N=3601)		Age >60 (Ref:<50) (N=2528)	
				OR	P-value	OR	P-value	OR	P-value
FAF1	1:50682573	rs17106184	A/G	0.70	[0.48,1.03]	1.01	[0.69, 1.50]	0.86	[0.48, 1.53]
NOTCH2	1:120319482	rs10923931	A/C	0.90	[0.68,1.20]	1.23	[0.91,1.66]	0.67	[0.41,1.11]
LINC00538 - PROX1	1:212225879	rs340874	G/A	1.24	[0.98,1.57]	1.11	[0.87,1.42]	1.15	[0.79,1.66]
GCKR	2:27594741	rs780094	A/G	1.08	[0.89,1.31]	0.97	[0.79,1.19]	1.13	[0.83,1.54]
THADA	2:43586327	rs7578597	G/A	0.66*	[0.44,0.98]	1.05	[0.70, 1.57]	1.51	[0.60, 3.76]
GRB14-COBLL1	2:165210095	rs13389219	A/G	0.92	[0.70,1.20]	1.21	[0.91,1.63]	1.20	[0.76,1.92]
LOC646736	2:226801989	rs7578326	G/A	0.98	[0.74,1.29]	0.96	[0.72,1.28]	1.01	[0.63,1.60]
PPARG	3:12264800	rs1801282	G/C	0.93	[0.61,1.41]	1.12	[0.72,1.73]	0.73	[0.37,1.43]
ADAMTS9-AS2	3:64686944	rs4607103	A/G	0.96	[0.75,1.23]	0.86	[0.66,1.12]	0.82	[0.56,1.21]
ADCY5	3:124548468	rs11708067	G/A	0.85	[0.60,1.21]	0.92	[0.64,1.33]	0.95	[0.51,1.76]
SLC2A2	3:172200215	rs11920090	A/T	1.21	[0.87,1.69]	1.66**	[1.15,2.40]	0.94	[0.51,1.73]
IGF2BP2	3:186994381	rs4402960	A/C	1.23*	[1.01,1.49]	0.90	[0.73,1.10]	0.69*	[0.50,0.96]
LPP	3:189223217	rs6808574	A/G	0.57**	[0.38,0.85]	0.89	[0.56,1.39]	0.61*	[0.37,0.99]
WFS1	4:6353923	rs1801214	G/A	0.86	[0.66,1.11]	1.01	[0.77,1.31]	0.75	[0.48,1.15]
TMEM154	4:153739925	rs6813195	A/G	1.08	[0.89,1.30]	0.92	[0.75,1.12]	0.90	[0.66,1.23]
ARL15	5:53307177	rs702634	G/A	0.91	[0.72,1.15]	0.88	[0.69,1.12]	1.24	[0.82,1.86]
ANKRD55-MAP3K1	5:55842508	rs459193	G/A	1.04	[0.75,1.45]	1.02	[0.72,1.43]	0.84	[0.48,1.47]
ZBED3-AS1	5:76460705	rs4457053	G/A	1.11	[0.84,1.46]	1.00	[0.75,1.34]	0.64	[0.41,1.01]
SSR1-RREB1	6:3672354	rs9505118	G/A	0.89	[0.74,1.08]	1.05	[0.86,1.29]	1.01	[0.73,1.38]
CDKAL1	6:20811931	rs6931514	G/A	1.03	[0.85,1.25]	1.03	[0.84,1.25]	1.18	[0.86,1.61]
POU5F1-TCF19	6:31244432	rs3130501	A/G	0.94	[0.76,1.15]	1.04	[0.84,1.29]	1.17	[0.84,1.63]
VEGFA - C6orf223	6:43919740	rs9472138	A/G	0.99	[0.76,1.28]	0.91	[0.70,1.19]	0.59*	[0.39,0.90]
DGKB - AGMO	7:15030834	rs2191349	C/A	0.89	[0.73,1.10]	0.97	[0.79,1.20]	0.91	[0.66,1.27]
JAZF1	7:28147081	rs864745	G/A	1.05	[0.84,1.31]	1.15	[0.91,1.45]	0.77	[0.54,1.10]

GCK - YKT6	7:44202193	rs4607517	A/G	0.70*	[0.53,0.93]	0.87	[0.65,1.16]	1.16	[0.73,1.86]
ACHE	7:100328013	rs7636	A/G	0.78	[0.41,1.47]	1.30	[0.66,2.55]	1.41	[0.46,4.35]
FSCN3 - PAX4	7:127034139	rs10229583	A/G	0.90	[0.72,1.11]	0.83	[0.66,1.03]	0.70*	[0.50,0.99]
KLF14 - MIR29A	7:130117394	rs972283	A/G	1.00	[0.82,1.21]	0.95	[0.77,1.16]	0.81	[0.58,1.12]
ANK1	8:41638405	rs516946	A/G	1.01	[0.76,1.32]	1.11	[0.83,1.49]	1.02	[0.66,1.57]
TP53INP1	8:96029687	rs896854	A/G	1.17	[0.96,1.43]	0.86	[0.69,1.05]	1.02	[0.74,1.42]
SLC30A8	8:118254206	rs3802177	A/G	0.93	[0.77,1.13]	0.84	[0.69,1.03]	1.10	[0.80,1.50]
GLIS3	9:4277466	rs7041847	A/G	0.69**	[0.54,0.88]	1.07	[0.83, 1.38]	1.26	[0.87,1.84]
CDKN2B-AS1 - DMRTA1	9:22123284	rs10965250	A/G	0.94	[0.77,1.15]	1.00	[0.81,1.23]	0.91	[0.66,1.25]
TLE1-FAM75D5	9:83498768	rs2796441	G/A	0.89	[0.70,1.13]	0.78	[0.61,1.00]	0.95	[0.65,1.39]
CDC123 - MIR4480	10:12347900	rs11257655	A/G	1.28*	[1.02,1.62]	1.09	[0.85,1.39]	1.16	[0.76,1.77]
VPS26A	10:70601480	rs1802295	A/G	0.95	[0.73,1.23]	0.80	[0.61,1.04]	0.86	[0.56,1.35]
ZMIZ1	10:80612637	rs12571751	G/A	0.89	[0.73,1.07]	1.17	[0.95,1.43]	1.10	[0.81,1.49]
IDE - RPL11P4	10:94452862	rs6583826	G/A	1.13	[0.91,1.40]	0.95	[0.76,1.18]	0.88	[0.62,1.25]
HHEX - EXOC6	10:94337810	rs1111875	G/A	1.19	[0.97,1.45]	0.87	[0.70,1.07]	0.92	[0.67,1.28]
ADRA2A - BTBD7P2	10:113032083	rs10885122	A/C	0.86	[0.66,1.11]	0.95	[0.73,1.25]	0.51**	[0.33,0.78]
TCF7L2	10:114748339	rs7903146	A/G	1.01	[0.78,1.30]	0.98	[0.74,1.28]	0.60*	[0.39,0.92]
KCNQ1; KCNQ1OT1	11:2648047	rs231362	A/G	1.08	[0.84,1.39]	0.95	[0.73,1.24]	0.97	[0.63,1.50]
KCNJ11	11:17365206	rs5215	G/A	1.00	[0.82,1.21]	0.79*	[0.64,0.97]	1.10	[0.79,1.52]
MADD	11:47292896	rs7944584	T/A	0.91	[0.66,1.27]	1.16	[0.82,1.64]	1.12	[0.66,1.89]
FADS1	11:61328054	rs174550	A/G	1.37*	[1.07,1.76]	1.02	[0.79,1.33]	0.93	[0.62,1.37]
ARAP1	11:72110746	rs1552224	C/A	0.78	[0.58,1.04]	1.08	[0.79,1.47]	1.58	[0.94,2.67]
FAT3 - MTNR1B	11:92313476	rs1387153	A/G	0.99	[0.82,1.20]	0.94	[0.77,1.15]	1.05	[0.77,1.43]
KLHDC5	12:27856417	rs10842994	A/G	1.02	[0.80,1.30]	0.75*	[0.59,0.97]	0.93	[0.63,1.36]
IGF1	12:101399699	rs35767	A/G	0.97	[0.79,1.19]	0.91	[0.74,1.13]	0.90	[0.65,1.25]
RPL12P33 - HNF1A-AS1	12:122013881	rs7305618	G/A	1.03	[0.81,1.30]	1.01	[0.79,1.30]	1.08	[0.75,1.55]
MPHOSPH9	12:119887315	rs4275659	A/G	1.00	[0.82,1.21]	1.00	[0.81,1.22]	1.00	[0.72,1.39]
NDFIP2 - SPRY2	13:79615157	rs1359790	A/G	0.98	[0.77,1.24]	1.07	[0.84,1.38]	1.19	[0.81,1.75]
C2CD4A - C2CD4B	15:60201306	rs1436953	A/G	0.88	[0.73,1.07]	1.15	[0.94,1.40]	0.82	[0.60,1.12]
HMG20A-LINGO1	15:75619817	rs7177055	A/G	0.82*	[0.68,0.99]	0.97	[0.80,1.18]	1.00	[0.71,1.39]

ZFAND6 - FAH	15:78219277	rs11634397	G/A	1.36	[0.99,1.87]	0.98	[0.71,1.37]	1.20	[0.68,2.13]
PRC1; LOC100507118	15:89322341	rs8042680	C/A	0.79	[0.58,1.08]	1.15	[0.83,1.61]	1.45	[0.82,2.56]
FTO	16:52378028	rs9939609	A/T	1.20	[0.97,1.48]	1.06	[0.85,1.32]	0.95	[0.66,1.38]
CTRB2-CTRB1	16:73804746	rs7202877	C/A	1.07	[0.83,1.39]	0.76*	[0.58,1.00]	1.12	[0.73,1.73]
SRR	17:2163008	rs391300	A/G	1.16	[0.92,1.47]	1.17	[0.92,1.49]	0.99	[0.68,1.45]
MC4R	18:56035730	rs12970134	A/G	0.98	[0.78,1.23]	1.13	[0.89,1.43]	1.14	[0.75,1.73]
PEPD	19:38584848	rs3786897	G/A	0.90	[0.74,1.09]	1.06	[0.86,1.30]	1.04	[0.75,1.45]
HNF4A	20:42422681	rs4812829	A/G	0.97	[0.80,1.17]	0.77*	[0.63,0.95]	0.90	[0.66,1.23]

Denotes statistically significant at *P<0.05; ** P<0.01; ***P<0.001

CHAPTER 6: DISCUSSION

6.1 General Discussion

T2D prevalence is increasing globally and particularly rapidly in Asian countries. T2D has been relatively understudied in the Malaysian population, in spite of Malaysia having one of the highest comparative prevalences of T2D among Asian nations. Malaysia is a multi-ethnic society comprising three major ancestral groups: Malays, Chinese and Indians. The prevalence of T2D in Malaysia differs between these ancestral groups, being highest in Indians, intermediate in Malays and lowest in Chinese; these differences exist in spite of these ancestral groups sharing a similar environment. The factors driving the T2D epidemic and variation among population prevalences of T2D in Malaysia remain unclear. To provide knowledge with potential public health relevance for managing T2D in Malaysia, this thesis sought to investigate the role and relative contributions of genetic and environmental risk factors, together with gene-environment interactions, to T2D in Malaysian population groups.

This research was divided into four components accordingly to the research objectives. The first (Chapter 2) comprised a comprehensive literature review. The second (Chapter 3) sought to assess the effects of known T2D genetic risk factors in the Malaysian population. The third (Chapter 4) assessed the association and contribution of environmental risk factors to T2D in Malaysia, while the fourth (Chapter 5) assessed gene-environment interactions. Each set of research questions were investigated both within the three major Malaysian population groups, and across the three groups combined.

6.1.1 Comprehensive Literature Review

Chapter 2 of this thesis comprised a comprehensive review of published T2D risk factors, both non-genetic and genetic. Genetic findings included results from GWAS published from 2007 [15] to 2013 [108], the most recent study at the time of writing [9].This review article included the compilation and description of 118 genetic risk variants found to be significantly associated with T2D in various populations. These risk variants provided the base set for selecting genetic variants to be studied in the remaining Chapters in the Malaysian sample.

Chapter 2 also discussed known population differences in T2D, in relation to prevalence, risk factor profiles and genetic risk alleles potentially contributing to the escalating prevalence of T2D in Asia. Studies discussed in this review highlighted the utility of T2D genetic and non-genetic research in diverse, multi-ethnic populations, to enhance risk factor identification and reveal factors potentially underlying population risk differences.

6.1.2 Genetic Risk Study

The genetic study reported in Chapter 3 represents the first detailed genetic study of T2D in Malaysia. This study involved the selection and analysis of 62 independent SNPs, each of which has shown compelling association with T2D in one or more previous large-scale studies. Prior GWAS of T2D, which informed the SNP selection, were compiled in Chapter 2. This genetic study identified seven out of 62 SNPs showing multiplicity-adjusted significant association ($P< 8.06 \times 10^{-4}$) with T2D in meta-analysis of the three ancestral groups. These SNPs were located in genes *TCF7L2*, *CDKN2A*, *FTO*, *PPARG*, *GCK*, *MC4R* and *ADCY5*. In addition, 10 further variants reached nominal significance at $P<0.05$; being rs1801214 in *WFS1* ($P=5\times10^{-3}$), rs6931514 in *CDKAL1* ($P=2\times10^{-3}$), rs3802177 in *SLC30A8* ($P= 7\times10^{-3}$), rs2796441 in *TLE1-FAM75D5* ($P=0.03$), rs1111875 in *HHEX - EXOC6* ($P= 1\times10^{-3}$), rs6583826 in *IDE - RPL11P4* ($P=0.02$), rs174550 in *FADS1* ($P=1\times10^{-3}$), rs1552224 in *ARAP1* ($P=0.01$), rs7177055 in *HMG20A-LINGO1* ($P=0.02$) and rs8042680 in *PRC1; LOC100507118* ($P=0.04$). Each of these genes is involved in a biological pathway influencing diabetes pathophysiology, including pancreatic beta-cell development/function, insulin availability, glucose utilisation, fatty acid concentrations and obesity. Of note, *PPARG* is a known target for thiazolidinediones, antidiabetic drugs which have been shown to improve insulin sensitivity and to reduce plasma glucose and blood pressure in persons with T2D [109]. This research thus confirmed the involvement of these pathophysiologically relevant SNPs in T2D in the Malaysian population.

This research was however, unable to confirm association for a range of individual genetic variants previously associated with T2D. In part, this likely reflected insufficient statistical power to identify variants with small individual effect. In this study, the sample was relatively small (N=4077) compared to the samples aggregated by international consortia (N=110, 452) [110]. Initially published effect sizes (odds ratios) for many variants were small, generally ranging from 1.0 to 1.2 [111]. Even smaller effect sizes were reported in subsequent studies for a range of variants, including similar populations from Singapore [14, 91]. This may be due to the phenomenon known as “winner’s curse”, or upward bias of effect estimates in the initial study [112]. Smaller effect sizes may also reflect lower linkage disequilibrium between assessed and underlying causal variants in the subsequently studied population, compared to European populations often used in the original study. Attenuated effect sizes or linkage disequilibrium will each diminish power to detect trait-variant association. Based on our sample size, we had 38%, 72% and 85-89% power to identify risk alleles with frequency 0.1, 0.2 and 0.3-0.5 respectively for a genetic risk ratio of 1.2. For a true risk ratio of 1.1, power was low, ranging from 4% to 19% across allele frequencies. Power of our study may have been further reduced by modest sample sizes in the individual ancestral subgroups, and heterogeneity of effect sizes and/or allele frequencies between subgroups.

Despite restricted power for testing individual variants, QQ-plots revealed an excess of nominally associated variants compared to chance expectation. Formal tests also showed a significantly elevated number of SNPs (28 out of 56 SNPs) whose estimated effect direction was consistent with

earlier studies (50% versus12.5%; binomial P=9.97 x 10^{-9}). This suggests that many of the assessed SNPs may well influence T2D risk in the Malaysian population and may well demonstrate more significant association in a larger sample.

It is known that T2D has a substantial genetic component with heritability estimates on the order of 30-70% [82, 113] but to date most reported variants individually account for only a small proportion of T2D heritability [71]. In parallel, this study indicated that the GRS of 62 variants explained less than 2% of overall T2D risk in any individual group, being highest in Chinese (1.7%), intermediate in Malay (1.6%) and lowest in Indians (1.0%). Thus, the 62 common variants assessed account for only a small proportion of overall T2D heritability in the Malaysian population, with a substantial proportion of heritability remaining unaccounted for. Since a relatively large set of common genetic variants appears to account for only a small proportion of T2D risk, the rapidly escalating T2D prevalence in Malaysia may be unlikely to result solely from common genetic variants. Further, since these well-established common variants each had small individual estimated effects, individual common risk variants for T2D would appear to provide limited utility for clinical risk prediction.

6.1.3 Environmental Risk Study

Findings from the first part of this research suggested that known, common T2D genetic risk variants account for only a small proportion of T2D risk in Malaysia. Expanding on this finding, the next stage of this research investigated the association and relative contribution of environmental (non-genetic) risk factors, to assess whether these contribute more to T2D risk in Malaysian populations than genetic factors alone. In research comprising Chapter 4 of this thesis, it was found that a combination of four non-genetic risk factors: age, gender, waist to hip ratio (WHR) and physical inactivity, accounted for about 20% of T2D risk in the combined Malaysian sample. The combination of these four environmental risks alone accounted for an estimated 26% of risk variation in Chinese, 18.5% in Indians and 15.1% in Malays. This suggests that rather than genetic risk factors, major contributors to the increasing T2D prevalence in Malaysia are determinants of obesity such as diet and physical inactivity, together with the ageing population. It is well established that reduction in physical activity increased the risk of T2D, irrespective of ancestry [114, 115]. It is also known that higher intakes of refined carbohydrates, saturated fats, and trans fats can increase T2D risk, while diets characterised by a low glycemic index and high dietary fibre intake are associated with decreased risk of T2D [116, 117]. Thus, these findings were broadly consistent with those from other ancestral groups.

In addition to variance explained, predictive accuracy of the four risk factors was quantified using the Area Under the Receiver Operator Characteristic Curve (AUC). These values showed a similar gradient across the respective ancestral groups, ranging from AUC=0.75 in Indians, AUC=0.78 in Malays and AUC=0.83 in Chinese, with the estimate in Chinese being significantly higher than the Indian value. These differences largely resulted from differential effects of waist-to-hip ratio (WHR) across the groups,

which may reflect ancestral differences in body fat percentage and the risk associated with abdominal obesity. WHR is an established indicator of insulin resistance and metabolic disease risk, and it is widely agreed that visceral fat percentage, rather than subcutaneous fat percentage, correlates more to risk of metabolic diseases such as T2D [118-120]. Abdominal obesity increases the risk of T2D by increasing the secretion of non-esterified fatty acids and adipocytokines such as tumour necrosis factor-α, and reducing adiponectin, leading to insulin resistance and T2D [121-124].

Asian individuals have been found to have a higher distribution of body fat around organs and in the abdominal area with concomitantly lower muscle mass, compared to Europeans with the same healthy BMI or WC [125]. Among the Malaysian population groups, body fat percentage tends to be naturally highest in Indians, followed respectively by Malays and Chinese. This finding suggests that for a Chinese person with a given level of abdominal obesity, the risk of developing T2D is increased disproportionately compared to the other two ancestral groups. A similar finding was observed in Chinese participants in the Multi-Ethnic Study of Atherosclerosis [126], which also found evidence for a greater risk of diabetes resulting from obesity in Chinese individuals, compared to those of European, Hispanic and African ancestry. This finding has potential significance for targeted public health interventions for T2D in Chinese Malaysians. More detailed studies focussed on visceral fat levels among the Chinese, Malay and Indian populations may shed light on its respective contributions to disease prevalence in different ancestral groups. In addition, the development of ethnicity-specific anthropometric measurement cut-points may be warranted, in order to identify culturally specific predictors of obesity and disease risk.

With respect to the association of age with T2D in this research, the risk of T2D is known to increase with advancing age due to age-related reductions in skeletal muscle mass (sarcopenia) and activation of glycogen synthase, and increases in visceral adiposity, leading to insulin resistance and glucose intolerance [127-129].

Although diet is likely an important contributor to abdominal obesity, this study was unable to detect significant association between high intakes of trans-fat or coffee consumption with T2D. High intake of trans-fat measured by deep frying, sautéing and use of coconut milk has previously been associated with increased cardiometabolic risk and insulin resistance [86, 130] while coffee consumption has been associated with improved insulin sensitivity and pancreatic beta-cell function [83, 131]. Possible explanations for these null associations are measurement error and recall bias, causing data inaccuracies that are common in nutritional epidemiological research, and leading to a loss of power to detect true nutrient intake. Typical cohort studies, including this study, measured diet using food frequency questionnaires, which suffer from measurement errors. Error in measuring nutritional intake can be considerable, and compounded by daily and seasonal variation in an individual's diet. Several important dietary factors were also not measured in this study, including polished white rice and refined wheat, which are staple foods in Asia [132], sugar-sweetened beverages [133] and Western-style fast food [134].

Inability to assess these important dietary variables – some or all of which may contribute to T2D variation in Malaysia – is a limitation of this research.

6.1.4 Gene-Environment Interaction

Although genetic and environmental (non-genetic) risk factors respectively contributed about 2 and 20% of disease risk in this study, respectively, there remains a substantial component of risk left unexplained. Some previous T2D research has highlighted the importance of risk factor combinations, and provided evidence of interactions between genetic and environmental factors in mediating disease risk [135-137]. However, studies of gene-environment interaction in T2D in the Asian context are limited, and few if any studies have specifically targeted Malaysians. Interactions between genetic and the environmental risk factors in the Malaysian context, and specifically among the individual Malay, Indian, and Chinese ancestral groups, may assist in explaining the missing heritability for T2D.

In the gene-environment interaction study reported in Chapter 5, both multiplicative and additive interactions were assessed. We found no significant evidence of gene-environment interaction, either for individual SNPs or a Genetic Risk Score (GRS). One likely explanation for these results is the large samples required to detect gene-environment interactions [76, 77]; this study limitation is discussed further below.

Interestingly, we found some evidence of gradient of GRS effects across ordered strata of BMI, with the GRS having progressively larger effects with decreasing levels of BMI. In addition, nominally significant gene-environment interaction ($P<0.05$) was observed for 33 individual SNP-environmental combinations, with many of these involving BMI; thus true gene-environment interactions for BMI may exist. A similar gradient of BMI effects was found in the recent EPIC InterAct Case-Cohort study including nine European samples. This large study also found a larger size of genetic risk score effects for T2D in participants who were leaner, both in terms of BMI (Hazard Ratio (HR): 1.6 in normal, 1.46 in overweight and 1.27 in obese) and waist circumference (HR: 1.6 in low WC, 1.53 in medium WC and 1.29 in high WC) [138]. Another European GWAS meta-analysis stratified by BMI identified a novel variant in the *LAMA1* gene, which only showed association with T2D in lean cases (OR =1.13 in lean cases compared to OR =1.03 in obese cases) [139]. Further, this same study found that 29 of 36 known T2D associated variants had a higher effect estimate (OR) for T2D in lean, relative to obese cases. Similarly, the Slim Initiative in Genomic Medicine for the Americas (SIGMA) consortium identified stronger T2D association for five SNP near *SLC16A11* among younger and leaner individuals of Mexican and Latin American descent [44]. Taken together, these findings suggest that lean T2D cases may have a higher relative contribution of genetic risk factors to their disease, compared to overweight or obese cases.

Interestingly, previous studies have reported up to a five-fold higher prevalence of lean T2D among Asian, compared to European, populations [10, 140, 141]. Similar results have been reported in other underdeveloped countries, e.g., see [140]. In this study, 5 out of 17 T2D variants (*TCF7L2*, *CDKN2A*, *HHEX*, *CDKAL1* and *HMG20A-LINGO1*) showing significant or nominally significant association with T2D in the Malaysian population were previously associated with lean T2D [10, 139]. Such effects may contribute to the observed effect gradient of the GRS across strata of BMI in this Malaysian sample, in accordance with other findings that Asians have a higher prevalence of lean T2D. Epidemiological studies have indicated that lean T2D patients tend to develop rapid beta cell failure in the condition of impaired insulin sensitivity [5, 142]. It has been suggested that in susceptible populations, beta cell development and function could be impaired by poor foetal and infant growth resulting from intrauterine nutritional deprivation [140]. Such influences from early life could contribute to later disease in the adult, in the absence of obesity.

For these and additional reasons, the utility of BMI as a marker of T2D risk may be reduced in some Asian, compared to European populations. In studies of some Asian populations, the risk of T2D has been shown to increase at comparatively normal BMI in Asian populations, suggesting BMI may not be a good measure of actual cardiometabolic risk [5]. Alternatively, WHR has been found to be superior as a predictor of T2D risk, especially in Asian populations [118-120]. A previous study conducted in a multiethnic population highlighted that Asians have greater adiposity or visceral fat than their European counterparts for a given BMI or waist circumference [143-146]. Visceral adiposity increases fatty acid influx to the liver, which alters adipokine production, leading to a fatty liver and hepatic insulin resistance [147].

In our study, the studied environmental risk factors explained a higher proportion of risk in Chinese, compared to Malays and Indians, based on AUC estimates. These findings, however, do not correspond with the gradient of prevalence of T2D in Malaysia, which is highest in Indians, followed by Malays and Chinese. The elevated prevalence of T2D in Indians may be due to other factors not accounted for in this study, and warrant further investigation. Proposed explanations include the thrifty genotype and thrifty phenotype hypotheses.

The thrifty genotype hypothesis postulates a mismatch between ancestral genes and the modern environment [57]. This hypothesis relates to populations that have experienced historical nutritional deprivation, characterised by alternating periods of feast and famine. During periods of famine, natural selection is proposed to have favoured individuals carrying so-called “thrifty” alleles promoting the storage of fat and energy. However, in modern civilisations increasingly exposed to “Westernised” diets, those with such “thrifty” genetic alleles may have a higher predisposition to obesity and related traits such as T2D. Such a phenomenon has been suggested for the Pima Indians of Arizona, who have a higher prevalence of T2D compared to white Americans [148].

In contrast, the thrifty phenotype hypothesis postulates a mismatch between intrauterine and adult life environments [57]. Similar conditions apply under this hypothesis, where poor foetal and infant growth due to nutritional deprivation during gestation/childhood causes an inability to effectively process high energy intake resulting from over-nutrition during adult life, contributing to detrimental health and the development of T2D and other chronic diseases. This biological response may be explained by impaired growth of the Beta cells and the islets of Langerhans during undernourishment, when there is a reduced requirement for insulin production [149]. However, a sudden change to high or over-nutrition then exposes the reduced capacity of Beta-cell function and may increase T2D risk. These hypotheses, while providing context for our research questions and results, were not directly tested in this study.

In summary, the current research considered the respective and combined contributions of genetic and environmental risk factors to T2D, as well as potential interactions between them. At the conclusion of this research, important questions remain. In particular, a substantial proportion of T2D heritability remains unexplained, by this or any other T2D research. The missing heritability phenomenon might reflect a number of factors influencing the genetic architecture of T2D, including polygenic inheritance (the existence of numerous risk variants with individually small effects), the presence of rare risk alleles, epistasis and possible epigenetic effects.

6.2 Missing Heritability

There are a number of possible explanations for the proportion of T2D heritability remaining unexplained (or “missing”) by numerous large-scale collaborative genetic studies. A major likely contributor is the existence of a large number of common risk variants with very small individual effects, precluding their identification in samples of available size [150, 151]. In addition, if the effects of some risk variants depend on genotypes of other variants, a component of heritability may be explained by the interaction between a combination of causal variants [152]. Epigenetic modifications, such as DNA methylation and histone modification may also account for a fraction of the “missing heritability”, by altering gene expression in heritable manner without affecting the underlying genomic sequence [153]. It has been shown that both the maternal environment during gestation, and the environment experienced during early infancy, can produce epigenetic changes influencing disease risk, independent of inherited variations in the DNA sequence [154]. Each of these possible factors is discussed below.

6.2.1 Polygenic Inheritance of Type 2 Diabetes

GWAS results suggest that T2D may be more genetically heterogeneous and polygenic than previously believed [150]. In addition to T2D, other complex diseases such as schizophrenia, bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, and type I diabetes have also been revealed as highly polygenic, with some sharing specific risk variants, a phenomenon known as pleiotropy [155, 156]. A comprehensive genetic analysis by Stahl and colleagues [151] revealed T2D as a highly polygenic disorder, with a combination of thousands of common SNPs explaining at least one-third of the total variation in T2D liability. For T2D, as well as MI/CAD and celiac disease, at least ~70% of their respective heritability was attributable to the additive effects of common SNPs [151].

Polygenic risk scores have also been shown to have utility for predicting T2D incidence. A study conducted by Vassy et al. [157] found that a polygenic score consisting of 62 known T2D-associated loci was significantly associated with incident T2D during 25 years of observation among young and middle-aged people of African and European ancestry. A similar study carried out by Vaxillaire et al. [158] found that polygenic scores incorporating information from 65 previously associated T2D SNPs was associated with increased incidence of impaired fasting glucose (IFG) and type 2 diabetes over 9 years of follow-up. These findings suggest that polygenic scores combining information across numerous risk alleles may have some utility for prospective assessment of T2D risk.

6.2.2 Rare Variants

The effects of rare genetic variants have been proposed as potentially explaining a component of risk variation, due to their potential to have larger effect sizes than those observed for common variants [159]. The frequency of any single rare (MAF<5%) or low-frequency variant is low (0.5% % MAF < 5%), but due to their large number in the human genome, they collectively represent a relatively common genetic phenomenon. In contrast to the common disease common variant (CDCV) hypothesis, the multiple rare variant (MRV) hypothesis asserts that there are many, large effect rare variants in the population and cases of a common, inherited disease may reflect the summation of the effects of a few of these moderate to high penetrance MRVs [160].

Recent advances in next-generation sequencing technology have allowed the rare variant hypothesis to be tested. A large, whole genome sequencing (WGS) study conducted in Icelanders, Danish and Iranian populations reported associations between rare variants in the *PDX1* and *PAM* genes and increased risk of T2D, and a protective effect of a rare variant in *CCND2* [161]. An alternative study using exome sequencing detected three novel, rare variants associated with fasting proinsulin or the insulinogenic index *TBC1D30*, *KANK1*, and *PAM* using 9,660 Finnish samples [162]. However, although these and other studies have detected association between rare variants and T2D, the effect sizes and

variance explained have been modest, and not supportive of the multiple rare variant (MRV) hypothesis [163].

While rare variants with large effect size may not importantly contribute to T2D heritability overall, they may be revealed as ethnicity-specific variants [164]. A whole-exome sequencing study (WES) using 3,756 Latino participants detected association of an ethnicity-specific rare missense variant in *HNF1A* (c.1522G > A [p.E508K]) and T2D, with a large effect size (OR, 4.96; 95% CI, 2.93–8.38) [165]. This finding highlights the value of studying diverse populations, underscoring the potential value of the present Malaysian sample for future genetic studies. Historically, progress in human disease genetics has been made by studying unique populations (e.g., isolates or diverse ethnic groups) or studying families selected on the basis of segregating specific, relevant phenotypes. Both unique populations and selected family groups may harbour unique rare variants and provide insight into disease pathogenesis.

6.2.3 Epistasis or Gene-gene Interaction

Epistasis or gene-gene interaction has been proposed as contributing to the missing heritability for complex diseases [166]. “Phantom heritability” created by genetic interactions may result in underestimation of the variance explained by identified risk variants. Such effects may reduce the limit of genetic variance explained to below 100%, even when all variants affecting the trait have been discovered [167]. Several studies have provided evidence for epistatic effects in T2D. An early study reported interaction between *EGFR* and *RXRG* variants in relation to the progression of diabetic nephropathy among Han Chinese [168]. Another study conducted in a Chinese population reported interactions among variants in *RAS*-related genes [166], while a recent study suggested epistasis between variants in *TCF7L2* and *WNT2B* in relation to susceptibility in Han Chinese [169]. Alternatively, a study conducted in Northern European individuals reported significant interaction between variants at1q21-25 and 10q23-26 in relation to T2D [170].

Study power represents a major challenge for comprehensive studies of gene-gene interaction, since interactions of modest effect are difficult to detect without extremely large sample sizes. Many studies reporting such analyses have had insufficient statistical power to robustly evaluate a large number of gene-gene interactions, partly reflecting the multiplicity challenges arising from the large number of possible pairwise combinations between genetic variation measured genome-wide. Such analyses also face computational challenges, although advances in this area have been made. Early attempts to study epistasis in complex diseases often focused on a single marker, or interactions between candidate regions [171, 172]. However, recent GWAS studies show the feasibility of newer methods for analysing gene-gene interaction using novel statistical approaches, such as global tests and joint Bayesian analysis of subphenotypes and epistasis (JBASE) [173, 174]. Future studies of epistasis in T2D will benefit from

ongoing computational advances and increasing sample sizes. Such studies may produce insights into the contribution of epistatic effects to T2D risk and pathogenesis.

6.2.4 Epigenetic Modifications

Epigenetic modifications such as DNA methylation and histone modification can contribute to heritable changes in gene expression at the level of the cell, tissue, or organism, without affecting the underlying genomic sequence. Since the DNA sequence is not changed, this epigenetic contribution would be systematically missed by conventional DNA sequence-based analyses, and evade detection in conventional GWAS. A recent study of DNA methylation [175] found that genes of the DNA replication process group (with upregulation of the *apex1, mcm2, mcm4, orc3*, *lig1*, and *dnmt*1 genes) were altered in the diabetic state and these molecular changes continued into the state of metabolic memory. This study found DNA methylation changes as far as 6–13 kb upstream of the transcription start site for these genes, suggesting potential higher levels of epigenetic control and providing a potential explanation for the heritable nature of diabetic metabolic memory. A methylation study in islets of the pancreas identified 1,649 CpG methylation targets and 853 genes, including *TCF7L2*, *FTO* and *KCNQ1*, with differential DNA methylation in T2D islets [176]. Further functional analyses demonstrated that the identified genes affected pancreatic β- and α-cells, as Exoc3l gene silencing reduced exocytosis, and overexpression of Cdkn1a, Pde7b and Sept9 perturbed insulin and glucagon secretion in clonal β- and α-cells, respectively [177]. This finding highlighted the potential importance of epigenetic effects in the pathogenesis of T2D. Another recent study found 101,911 SNP-CpG pairs (mQTLs) in *cis* and 5,342 SNP-CpG pairs in *trans* showing significant associations between genotypes (in *DCY3/POMC*, *APOA5*, *CETP*, *FADS2*, *GCKR*, *SORT1* and *LEPR*) and DNA methylation in adipose tissue [178]. This study provided compelling evidence that genetic variants can mediate effects on metabolic traits, including lipid, homeostatic model assessment of insulin resistance (HOMA-IR) and glucose traits, via altered DNA methylation in human adipose tissue [178].

Recent studies have also found that DNA methylation in humans can be influenced by diet and exercise [179-181]. Jacobsen showed that short-term, high-fat overfeeding in healthy young men introduced widespread DNA methylation changes affecting 6,508 genes [179]. Conversely, aerobic exercise has been shown to attenuate epigenetic modifications at *PGC1* induced by high-energy diets and reduced physical activity, inhibiting T2D onset [182]. There is also evidence of population differences in methylation scores, where combined methylation scores for five T2D-associated loci (*ABCG1, PHOSPHO1, SOCS3, SREBF1* and *TXNIP*) were substantially higher among Asian Indians than Europeans [183]. The affected genes have been associated with BMI, waist circumference, insulin concentrations, glucose concentration and HOMA-IR. Taken together, these findings suggest that epigenetic modification has an important role in the pathogenesis of T2D, and that relevant epigenetic differences may exist between Asians and Europeans. Comprehensive assessment of DNA methylation

may thus help to explain the increased risk of T2D in particular Asian groups (e.g., Indians) and in Asians overall, compared to other populations.

Evidence from a large family-based study showed that once they have occurred, epigenetic modifications can be stable across cell divisions and inherited by offspring, manifesting as parent-of-origin effects such as larger effect sizes for T2D risk variants when the risk allele is transmitted by the mother, rather than the father [184]. Foetal growth development can also be influenced by epigenetic changes in response to nutritional intake, leading to permanent changes in glucose-insulin metabolism and increased chronic disease susceptibility as an adult [185]. According to the "metabolic memory" hypothesis, cells can memorize changes in glucose concentrations by inducing histone modifications in endothelial cells [186]. This hypothesis was supported by two large studies, the UK Prospective Diabetes Study (UKPDS) and Diabetes Control Complications Trial (DCCT) studies, which showed that initially good metabolic control was associated with reduced chances of diabetic complications decades later [186].

In addition to epigenetic changes, non-coding RNAs such as MicroRNAs (miRNAs) have recently emerged as important regulators of gene expression and function. Several studies have implicated miRNAs in T2D and inflammation, with common SNPs altering the sequence of miRNAs in several T2D susceptibility loci [187].

Other forms of non-coding RNAs, such as piRNAs (PIWI-interacting RNAs), snoRNAs (small nucleolar RNAs), lincRNAs (long intergenic non-coding RNAs), and lncRNAs (long non-coding RNAs), may also contribute to variation in gene expression influencing T2D susceptibility. miRNA manipulation has been explored as a novel therapeutic modality for reducing and preventing complication of T2D [188]. Further studies of these various genomic variations will help to elucidate their pathogenic roles and potential utility as T2D therapeutic targets.

6.3 Study Limitations

6.3.1. Statistical Power

A limitation of this research was the relatively modest sample size (N=4077) compared to other International consortia, e.g., (N=110, 452) [110]. This was likely an important factor underlying the inability to confirm a range of associations for genetic variants previously associated with T2D, reflecting insufficient statistical power to identify variants with small individual effect. Insufficient power was also a likely contributor to the negative gene-environment interaction tests. Not only was the genetic sample in our study relatively small, the effective sample size was reduced in studies of non-genetic risk factors due to one environmental factor of interest (physical activity) being only measured among a subset of GWAS participants, reflecting an upgrade to questionnaires during the research period.

It is known that identifying genetic main effects for complex disease requires substantial sample sizes, with even larger samples required to detect gene-environment interactions. Based on a broad, earlier recommendation by an NHGRI Expert Panel, cohorts of 200,000 to 500,000 participants may be required to identify important genetic effects for many common diseases such as T2D [189]. However, such large numbers of participants are costly and time-consuming to recruit and genotype, and can only be achieved by larger collaborative studies that combine the data collected for numerous, individual studies. Such collaborations will undoubtedly be necessary to identify many gene-environment interactions; however the requisite lifestyle (non-genetic) data may not be routinely collected, presenting an additional challenge.

Another approach to overcome sample size restrictions is to conduct meta-analyses of the results of different studies, although published literature for meta-analyses can be affected by publication bias. While meta-analysis of published and non-published data may help to resolve this issue, between-study heterogeneity in the way exposures and outcomes have been assessed and categorised is still challenging to resolve. Additional factors such as study design and measurement error in the variables also affect statistical power to detect gene-environment effects and should be considered in interpreting the results from these, and other gene-environment studies.

6.3.2 Bias and Measurement Error

Some measures of environmental exposure are challenging to accurately quantify and standardize, especially when based on self-report questionnaires. Measurement errors resulting from recall and reporting bias can occur when recorded data are highly dependent on participants' memory and subjective reporting. Recall bias occurs when disease status influences the collection and reporting of exposures; for example, questions about exposure to a putative cause might be asked many times of known cases but only once for those without disease. Reporting bias occurs when a participant provides answers in the direction they perceive are of interest, and avoids providing socially undesirable responses [190].

Measurement error is common when measuring nutrient intake, and likely affects the ability to detect association between nutritional factors and disease [191]. Typical cohort studies, including this study, measure diet using food frequency questionnaires, which are known to suffer measurement error not only due to recall and reporting bias, but also daily and seasonal variability of an individual's diet. Such measurement error might be a reason this study was unable to detect significant association between dietary factors and T2D.

Measurement error due to self-report may have also affected our measurement of physical activity, which did not show statistically significant association with T2D, in spite of having a known influence on T2D. Objective measurements of physical activity obtained via pedometers or

accelerometers provide more reliable measurements than self-report [192], with the correlation between self-reported physical activity and pedometer-assessed step count being potentially low [193]. The transition between two versions of physical activity questionnaires during the study might also influence the results. However, differences are likely to be small, given that difference between the original IPAQ and the modified IPAQ-M related only to language and cultural adaptation, based on translation from English to Malay [79].

Taken together, such measurement limitations may have caused over- or underreporting of environmental risk factors in this study. Such measurement error reduces statistical power. This was supported by a simulated gene-environment interaction study, which showed that even moderate decreases in measurement accuracy (correlation with true score of 0.4 vs 0.7) of environmental risk factors can result in a 20-fold reduction in statistical power to detect gene-environment interactions [194].

6.4 Clinical Implications of this research

Apart from shortcomings mentioned above, these findings may have some public health significance relating to the risk and burden of disease. Firstly, this research provides evidence potentially supporting targeted intervention strategies in population subgroups, suggesting a greater importance of abdominal obesity in Malaysians of Chinese ancestry. Alternately, this finding may reflect anthropometric differences between ancestral groups, suggesting the possible utility of ethnicity-specific anthropometric cut- points for evaluating T2D risk. Further, variation in genetic effects between cases in different obesity strata suggests the potential for personalised prevention recommendations for lean Asians with elevated genetic risk. In closing, although we found significant association between genetic risk variation and T2D, at this stage, the variance explained by common genetic variants appears too small to facilitate individual risk prediction.

6.5 Conclusions

Results of this research suggest that recent changes in lifestyle and diet contributing to abdominal obesity, combined with the ageing populations, are major contributors to the increased prevalence of T2D in Malaysia. The relative effect of common genetic risk variants appears small. Targeted public health interventions focussed on reducing obesity will help to mitigate the soaring prevalence of T2D in Malaysia. Introducing ethnicity-specific anthropometric guidelines may also be important for accurate disease risk estimation in diverse Malaysian population groups, given differences in body fat distribution among Asian groups, and between Asian and European reference groups.

CHAPTER 7: REFERENCES

1. Murea, M., L. Ma, and B.I. Freedman, *Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications.* Rev Diabet Stud, 2012. **9**(1): p. 6-22.
2. Habtewold, T.D., et al., *Comorbidity of depression and diabetes: an application of biopsychosocial model.* Int J Ment Health Syst, 2016. **10**: p. 74.
3. International Diabetes Federation, *IDF Diabetes Atlas 6th Ed.,*. 2015, International Diabetes Federation: Brussels, Belgium.
4. World Health Organization, *Global Report on Diabetes*. 2016, World Health Organization,: Geneva.
5. Ma, R.C. and J.C. Chan, *Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States.* Ann N Y Acad Sci, 2013. **1281**: p. 64-91.
6. International Diabetes Federation, *IDF Diabetes Atlas, 7th edn*. 2015, International Diabetes Federation: Brussels, Belgium.
7. Ramachandran, A., et al., *Trends in prevalence of diabetes in Asian countries.* World J Diabetes, 2012. **3**(6): p. 110-7.
8. Png, M.E., et al., *Current and future economic burden of diabetes among working-age adults in Asia: conservative estimates for Singapore from 2010-2050.* BMC Public Health, 2016. **16**: p. 153.
9. Abdullah N., A.J., Oldmeadow C., Scott R.J. & Holliday E.G, *The architecture of risk for type 2 diabetes: understanding Asia in the context of global findings.* International Journal of Endocrinology, 2014. **2014**: p. 21.
10. Kong, X., et al., *Genetic variants associated with lean and obese type 2 diabetes in a Han Chinese population: A case-control study.* Medicine (Baltimore), 2016. **95**(23): p. e3841.
11. Morris, A.P., et al., *Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes.* Nat Genet, 2012. **44**(9): p. 981-90.
12. Centers for Disease Control and Prevention, *National Diabetes Statistics Report Figures*. 2014, National Centers for Disease Control and Prevention: Atlanta, USA.
13. Hinney, A., C.I. Vogel, and J. Hebebrand, *From monogenic to polygenic obesity: recent advances.* Eur Child Adolesc Psychiatry, 2010. **19**(3): p. 297-310.
14. Sim, X., et al., *Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia.* PLoS Genet, 2011. **7**(4): p. e1001363.
15. Sladek, R., et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. **445**(7130): p. 881-5.
16. Qi, Q. and F.B. Hu, *Genetics of type 2 diabetes in European populations.* J Diabetes, 2012. **4**(3): p. 203-12.
17. Chimienti, F., et al., *Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules.* Diabetes, 2004. **53**(9): p. 2330-7.
18. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. **316**(5829): p. 1336-41.
19. Steinthorsdottir, V., et al., *A variant in CDKAL1 influences insulin response and risk of type 2 diabetes.* Nat Genet, 2007. **39**(6): p. 770-5.
20. Scott, L.J., et al., *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* Science, 2007. **316**(5829): p. 1341-5.
21. Diabetes Genetics Initiative of Broad Institute of, H., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.
22. Wellcome Trust Case Control, C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
23. Groenewoud, M.J., et al., *Variants of CDKAL1 and IGF2BP2 affect first-phase insulin secretion during hyperglycaemic clamps.* Diabetologia, 2008. **51**(9): p. 1659-63.
24. Pascoe, L., et al., *Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function.* Diabetes, 2007. **56**(12): p. 3101-4.
25. Dina, C., et al., *Variation in FTO contributes to childhood obesity and severe adult obesity.* Nat Genet, 2007. **39**(6): p. 724-6.
26. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. **316**(5826): p. 889-94.
27. Zeggini, E., et al., *Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.* Nat Genet, 2008. **40**(5): p. 638-45.

28. Yasuda, K., et al., *Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus.* Nat Genet, 2008. **40**(9): p. 1092-7.
29. Unoki, H., et al., *SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations.* Nat Genet, 2008. **40**(9): p. 1098-102.
30. Tsai, F.J., et al., *A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese.* PLoS Genet, 2010. **6**(2): p. e1000847.
31. Yamauchi, T., 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**(10): p. 864-8.
32. Shu, X.O., et al., *Identification of new genetic risk variants for type 2 diabetes.* PLoS Genet, 2010. **6**(9).
33. McDonough, C.W., et al., *A genome-wide association study for diabetic nephropathy genes in African Americans.* Kidney Int, 2011. **79**(5): p. 563-72.
34. Kooner, J.S., et al., *Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci.* Nat Genet, 2011. **43**(10): p. 984-9.
35. Below, J.E., et al., *Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals.* Diabetologia, 2011. **54**(8): p. 2047-55.
36. Parra, E.J., 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**(8): p. 2038-46.
37. Kwak, S.H. and K.S. Park, *Genetics of type 2 diabetes and potential clinical implications.* Arch Pharm Res, 2013. **36**(2): p. 167-77.
38. Cho, Y.S., et al., *Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians.* Nat Genet, 2012. **44**(1): p. 67-72.
39. Saxena, R., et al., *Genome-Wide Association Study Identifies a Novel Locus Contributing to Type 2 Diabetes Susceptibility in Sikhs of Punjabi Origin From India.* Diabetes, 2013.
40. Ma, R.C., et al., *Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4.* Diabetologia, 2013. **56**(6): p. 1291-305.
41. Al Safar, H.S., et al., *A Genome-Wide Search for Type 2 Diabetes Susceptibility Genes in an Extended Arab Family.* Ann Hum Genet, 2013.
42. Hara, K., et al., *Genome-wide association study identifies three novel loci for type 2 diabetes.* Hum Mol Genet, 2014. **23**(1): p. 239-46.
43. Hanson, R.L., et al., *A genome-wide association study in American Indians implicates DNER as a susceptibility locus for type 2 diabetes.* Diabetes, 2014. **63**(1): p. 369-76.
44. Consortium, S.T.D., et al., *Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico.* Nature, 2014. **506**(7486): p. 97-101.
45. Replication, D.I.G., et al., *Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility.* Nat Genet, 2014. **46**(3): p. 234-44.
46. Ma, R.C., et al., *Familial young-onset diabetes, pre-diabetes and cardiovascular disease are associated with genetic variants of DACH1 in Chinese.* PLoS One, 2014. **9**(1): p. e84770.
47. Ng, M.C., et al., *Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes.* PLoS Genet, 2014. **10**(8): p. e1004517.
48. Ghassibe-Sabbagh, M., et al., *T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility.* Sci Rep, 2014. **4**: p. 7351.
49. Anderson, D., et al., *First genome-wide association study in an Australian aboriginal population provides insights into genetic risk factors for body mass index and type 2 diabetes.* PLoS One, 2015. **10**(3): p. e0119333.
50. Imamura, M., et al., *Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes.* Nat Commun, 2016. **7**: p. 10531.
51. Cook, J.P. and A.P. Morris, *Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility.* Eur J Hum Genet, 2016. **24**(8): p. 1175-80.
52. Jorde, L.B., *Linkage disequilibrium and the search for complex disease genes.* Genome Res, 2000. **10**(10): p. 1435-44.
53. Rosenberg, N.A., et al., *Genome-wide association studies in diverse populations.* Nat Rev Genet, 2010. **11**(5): p. 356-66.

54. Ng, M.C., et al., *Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians.* Diabetes, 2008. **57**(8): p. 2226-33.
55. Chen, R., et al., *Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases.* PLoS Genet, 2012. **8**(4): p. e1002621.
56. Carulli, L., et al., *Review article: diabetes, genetics and ethnicity.* Aliment Pharmacol Ther, 2005. **22 Suppl 2**: p. 16-9.
57. Neel, J.V., *Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? 1962.* Bull World Health Organ, 1999. **77**(8): p. 694-703; discussion 692-3.
58. Hu, F.B., *Globalization of diabetes: the role of diet, lifestyle, and genes.* Diabetes Care, 2011. **34**(6): p. 1249-57.
59. Yeo, K.K., et al., *Ethnicity modifies the association between diabetes mellitus and ischaemic heart disease in Chinese, Malays and Asian Indians living in Singapore.* Diabetologia, 2006. **49**(12): p. 2866-73.
60. Sloan, N.R., *Ethnic distribution of diabetes mellitus in Hawaii.* JAMA, 1963. **183**: p. 419-24.
61. Tan, J.T., et al., *Polymorphisms identified through genome-wide association studies and their associations with type 2 diabetes in Chinese, Malays, and Asian-Indians in Singapore.* J Clin Endocrinol Metab, 2010. **95**(1): p. 390-7.
62. Chan, J.C., et al., *Diabetes in Asia: epidemiology, risk factors, and pathophysiology.* JAMA, 2009. **301**(20): p. 2129-40.
63. Williams, D.E., et al., *The effect of Indian or Anglo dietary preference on the incidence of diabetes in Pima Indians.* Diabetes Care, 2001. **24**(5): p. 811-6.
64. Smith, C.J., et al., *Survey of the diet of Pima Indians using quantitative food frequency assessment and 24-hour recall. Diabetic Renal Disease Study.* J Am Diet Assoc, 1996. **96**(8): p. 778-84.
65. Fujimoto, W.Y., et al., *Diabetes and diabetes risk factors in second- and third-generation Japanese Americans in Seattle, Washington.* Diabetes Res Clin Pract, 1994. **24 Suppl**: p. S43-52.
66. Qi, Q., et al., *Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies.* BMJ, 2014. **348**: p. g1610.
67. Qi, L., et al., *Genetic predisposition, Western dietary pattern, and the risk of type 2 diabetes in men.* Am J Clin Nutr, 2009. **89**(5): p. 1453-8.
68. Hindy, G., et al., *Role of TCF7L2 risk variant and dietary fibre intake on incident type 2 diabetes.* Diabetologia, 2012. **55**(10): p. 2646-2654.
69. Cornelis, M.C., et al., *TCF7L2, dietary carbohydrate, and risk of type 2 diabetes in US women.* Am J Clin Nutr, 2009. **89**(4): p. 1256-62.
70. Nair, A.K. and L.J. Baier, *Complex Genetics of Type 2 Diabetes and Effect Size: What have We Learned from Isolated Populations?* Rev Diabet Stud, 2015. **12**(3-4): p. 299-319.
71. Eichler, E.E., et al., *Missing heritability and strategies for finding the underlying causes of complex disease.* Nat Rev Genet, 2010. **11**(6): p. 446-50.
72. International Diabetes Federation, *IDF Diabetes Atlas Sixth Edition Poster Update 2014*. 2014.
73. Department of Statistics Malaysia, *Population Distribution and Basic Demographic Characteristics 2010*. 2011, Department of Statistics Malaysia,: Putrajaya, Malaysia.
74. Jamal R., S.Z., S.Z., Kamaruddin M.A., Jalal A.N, Ismail N., Kamil N.M, Abdullah N., Baharudin N., Hussin N.H, Othman H., Mahadi N. M., The Malaysian Cohort Group,, *Cohort profile: The Malaysian Cohort (TMC) project: a prospective study of non-communicable diseases in a multi-ethnic population.* The International Journal of Epidemiology, 2014: p. 9.
75. Gong, L., et al., *The FOXO1 Gene-Obesity Interaction Increases the Risk of Type 2 Diabetes Mellitus in a Chinese Han Population.* J Korean Med Sci, 2017. **32**(2): p. 264-271.
76. InterAct, C., *Investigation of gene-diet interactions in the incretin system and risk of type 2 diabetes: the EPIC-InterAct study.* Diabetologia, 2016. **59**(12): p. 2613-2621.
77. Xiao, S., et al., *Gene Polymorphism Association with Type 2 Diabetes and Related Gene-Gene and Gene-Environment Interactions in a Uyghur Population.* Med Sci Monit, 2016. **22**: p. 474-87.
78. Zheng, L. and Q. Li, *Impact of apolipoprotein E gene polymorphism and additional gene-obesity interaction on type 2 diabetes risk in a Chinese Han old population.* Obes Res Clin Pract, 2016.
79. Shamsuddin N Poh BK, S.Z.S.Z., Noor M.I & Jamal R, *Reliability and Validity of Malay Language Version of International Physical Activity Questionnaire (IPAQ-M) among the Malaysian Cohort Participants.* International Journal of Public Health Research, 2015. **5**(2): p. 643-53.

80. UK Biobank Coordinating Centre, *Protocol for a Large-Scale Prospective Epidemiological Resource*. 2007: Notthingham, UK.
81. Bjornland, T., et al., *Assessing gene-environment interaction effects of FTO, MC4R and lifestyle factors on obesity using an extreme phenotype sampling design: Results from the HUNT study.* PLoS One, 2017. **12**(4): p. e0175071.
82. Carlsson, S., et al., *Shared genetic influence of BMI, physical activity and type 2 diabetes: a twin study*. Diabetologia, 2013. **56**(5): p. 1031-5.
83. van Dam, R.M. and E.J. Feskens, *Coffee consumption and risk of type 2 diabetes mellitus.* Lancet, 2002. **360**(9344): p. 1477-8.
84. Wilson, P.W., et al., *Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study*. Arch Intern Med, 2007. **167**(10): p. 1068-74.
85. Cahill, L.E., et al., *Fried-food consumption and risk of type 2 diabetes and coronary artery disease: a prospective study in 2 cohorts of US women and men.* Am J Clin Nutr, 2014. **100**(2): p. 667-675.
86. Lopez-Garcia, E., et al., *Consumption of trans fatty acids is related to plasma biomarkers of inflammation and endothelial dysfunction.* J Nutr, 2005. **135**(3): p. 562-6.
87. Purcell S., *PLINK v1.07*. 2009.
88. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
89. Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A., *The History and Geography of Human Genes.* 1996, New Jersey: Princeton University Press.
90. Reich, D., A.L. Price, and N. Patterson, *Principal component analysis of genetic data*. Nat Genet, 2008. **40**(5): p. 491-492.
91. Teo, Y.Y., et al., *Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations.* Genome Res, 2009. **19**(11): p. 2154-62.
92. Hindorff LA, M.J., Morales J., Junkins HA., Hall PN., Klemm AK. & Manolio TA, *A Catalog of Published Genome-Wide Association Studies*, N.H.G.R. Institute, Editor. 2014, National HUman Genome Research Institute.
93. Johnson, A.D., et al., *SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap.* Bioinformatics, 2008. **24**(24): p. 2938-9.
94. Cooke Bailey, J.N. and R.P. Igo, *Genetic Risk Scores*, in *Current Protocols in Human Genetics*. 2001, John Wiley & Sons, Inc.
95. World Health Organization *Global Recommendations on Physical Activity for Health 18-64 years old*. 2011.
96. World Health Organization, *Obesity – Preventing and Managing the Global Epidemic: Report of a WHO Consultation on Obesity.* . 1998, World Health Oranization. p. 158.
97. World Health Organization *Obesity and overweight Factsheet*. 2013.
98. World Health Organization, *Obesity: Preventing and managing the global epidemic. Report of a WHO Consultation*, WHO, Editor. 2000, WHO,: Geneva. p. 16.
99. World Health Organization, *Waist circumference and Waist-Hip Ratio. Report of a WHO Expert Consultation*. 2008: Geneva. p. 47.
100. Azur, M.J., et al., *Multiple imputation by chained equations: what is it and how does it work?* Int J Methods Psychiatr Res, 2011. **20**(1): p. 40-9.
101. Graham, J.W., A.E. Olchowski, and T.D. Gilreath, *How many imputations are really needed? Some practical clarifications of multiple imputation theory.* Prev Sci, 2007. **8**(3): p. 206-13.
102. Graham, J.W., *Missing data analysis: making it work in the real world.* Annu Rev Psychol, 2009. **60**: p. 549-76.
103. Greenland S. Pearce N., *Modeling strategies in epidemiology: II. Basic alternatives.*, in *Modeling Startegy in Epidemiology*. 2014. p. 1-20.
104. Cleves., M.A., *From the help desk: Comparing areas under receiver operating characteristic curves from two or more probit or logit models.* the Stata Journal, 2002. **2**(3): p. 301-313.
105. DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson, *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.* Biometrics, 1988. **44**(3): p. 837-45.
106. VanderWeele, T.J., *On the distinction between interaction and effect modification.* Epidemiology, 2009. **20**(6): p. 863-71.
107. Knol, M.J. and T.J. VanderWeele, *Recommendations for presenting analyses of effect modification and interaction.* Int J Epidemiol, 2012. **41**(2): p. 514-20.
108. Pasquale, L.R., et al., *Exploring genome-wide - dietary heme iron intake interactions and the risk of type 2 diabetes.* Front Genet, 2013. **4**: p. 7.

109. Day, C., *Thiazolidinediones: a new class of antidiabetic drugs.* Diabet Med, 1999. **16**(3): p. 179-92.
110. Replication, D.I.G. and C. Meta-analysis, *Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility.* Nat Genet, 2014. **46**(3): p. 234-44.
111. Abdullah, N., et al., *Characterizing the genetic risk for Type 2 diabetes in a Malaysian multi-ethnic cohort.* Diabet Med, 2015.
112. Nakaoka, H. and I. Inoue, *The Winner's Curse*, in *eLS*. 2001, John Wiley & Sons, Ltd.
113. Poulsen, P., et al., *Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study.* Diabetologia, 1999. **42**(2): p. 139-45.
114. Laaksonen, D.E., et al., *Physical activity in the prevention of type 2 diabetes: the Finnish diabetes prevention study.* Diabetes, 2005. **54**(1): p. 158-65.
115. Mohan, V., et al., *Association of physical inactivity with components of metabolic syndrome and coronary artery disease--the Chennai Urban Population Study (CUPS no. 15).* Diabet Med, 2005. **22**(9): p. 1206-11.
116. Mohan, V., et al., *Dietary carbohydrates, glycaemic load, food groups and newly detected type 2 diabetes among urban Asian Indian population in Chennai, India (Chennai Urban Rural Epidemiology Study 59).* Br J Nutr, 2009. **102**(10): p. 1498-506.
117. Hu, F.B., R.M. van Dam, and S. Liu, *Diet and risk of Type II diabetes: the role of types of fat and carbohydrate.* Diabetologia, 2001. **44**(7): p. 805-17.
118. Kaur, P., et al., *A comparison of anthropometric indices for predicting hypertension and type 2 diabetes in a male industrial population of Chennai, South India.* Ethn Dis, 2008. **18**(1): p. 31-6.
119. Cheng, C.H., et al., *Waist-to-hip ratio is a better anthropometric index than body mass index for predicting the risk of type 2 diabetes in Taiwanese population.* Nutr Res, 2010. **30**(9): p. 585-93.
120. Xin, Z., et al., *Identifying obesity indicators which best correlate with type 2 diabetes in a Chinese population.* BMC Public Health, 2012. **12**: p. 732.
121. Bjorntorp, P., *Metabolic implications of body fat distribution.* Diabetes Care, 1991. **14**(12): p. 1132-43.
122. Bjorntorp, P., *"Portal" adipose tissue as a generator of risk factors for cardiovascular disease and diabetes.* Arteriosclerosis, 1990. **10**(4): p. 493-6.
123. Diez, J.J. and P. Iglesias, *The role of the novel adipocyte-derived hormone adiponectin in human disease.* Eur J Endocrinol, 2003. **148**(3): p. 293-300.
124. Despres, J.P. and I. Lemieux, *Abdominal obesity and metabolic syndrome.* Nature, 2006. **444**(7121): p. 881-7.
125. Lear, S.A., et al., *Visceral adipose tissue accumulation differs according to ethnic background: results of the Multicultural Community Health Assessment Trial (M-CHAT).* Am J Clin Nutr, 2007. **86**(2): p. 353-9.
126. Lutsey, P.L., et al., *Interactions between race/ethnicity and anthropometry in risk of incident diabetes: the multi-ethnic study of atherosclerosis.* Am J Epidemiol, 2010. **172**(2): p. 197-204.
127. Alberti, K.G., P. Zimmet, and J. Shaw, *International Diabetes Federation: a consensus on Type 2 diabetes prevention.* Diabet Med, 2007. **24**(5): p. 451-63.
128. Willey, K.A. and M.A. Singh, *Battling insulin resistance in elderly obese people with type 2 diabetes: bring on the heavy weights.* Diabetes Care, 2003. **26**(5): p. 1580-8.
129. Landi, F., G. Onder, and R. Bernabei, *Sarcopenia and diabetes: two sides of the same coin.* J Am Med Dir Assoc, 2013. **14**(8): p. 540-1.
130. Haag, M. and N.G. Dippenaar, *Dietary fats, fatty acids and insulin resistance: short review of a multifaceted connection.* Med Sci Monit, 2005. **11**(12): p. RA359-67.
131. Loopstra-Masters, R.C., et al., *Associations between the intake of caffeinated and decaffeinated coffee and measures of insulin sensitivity and beta cell function.* Diabetologia, 2011. **54**(2): p. 320-8.
132. Villegas, R., et al., *Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women.* Arch Intern Med, 2007. **167**(21): p. 2310-6.
133. Malik, V.S., M.B. Schulze, and F.B. Hu, *Intake of sugar-sweetened beverages and weight gain: a systematic review.* Am J Clin Nutr, 2006. **84**(2): p. 274-88.
134. Odegaard, A.O., et al., *Western-style fast food intake and cardiometabolic risk in an Eastern country.* Circulation, 2012. **126**(2): p. 182-8.
135. Franks, P.W. and G. Pare, *Putting the Genome in Context: Gene-Environment Interactions in Type 2 Diabetes.* Curr Diab Rep, 2016. **16**(7): p. 57.

136. Manning AK, *Gene-Environment Interaction: Methods and Examples in Type 2 Diabetes and Obesity.*, in *The Genetics of Type 2 Diabetes and Related Traits*. 2016, Springer. p. 259-73.
137. Samsom, M., et al., *Understanding the Importance of Gene and Environment in the Etiology and Prevention of Type 2 Diabetes Mellitus in High-Risk Populations.* Oral Health Case Rep, 2016. **2**(1).
138. Langenberg, C., et al., *Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study.* PLoS Med, 2014. **11**(5): p. e1001647.
139. Perry, J.R., 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**(5): p. e1002741.
140. Coleman, N.J., et al., *Lean versus obese diabetes mellitus patients in the United States minority population.* J Diabetes Complications, 2014. **28**(4): p. 500-5.
141. Whincup, P.H., et al., *Birth weight and risk of type 2 diabetes: a systematic review.* JAMA, 2008. **300**(24): p. 2886-97.
142. George, A.M., A.G. Jacob, and L. Fogelfeld, *Lean diabetes mellitus: An emerging entity in the era of obesity.* World J Diabetes, 2015. **6**(4): p. 613-20.
143. Deurenberg, P., M. Deurenberg-Yap, and S. Guricci, *Asians are different from Caucasians and from each other in their body mass index/body fat per cent relationship.* Obes Rev, 2002. **3**(3): p. 141-6.
144. Deurenberg-Yap, M., et al., *Body fat measurement among Singaporean Chinese, Malays and Indians: a comparative study using a four-compartment model and different two-compartment models.* Br J Nutr, 2001. **85**(4): p. 491-8.
145. Gurrici, S., et al., *Relationship between body fat and body mass index: differences between Indonesians and Dutch Caucasians.* Eur J Clin Nutr, 1998. **52**(11): p. 779-83.
146. Lear, S.A., et al., *The use of BMI and waist circumference as surrogates of body fat differs by ethnicity.* Obesity (Silver Spring), 2007. **15**(11): p. 2817-24.
147. Jung, U.J. and M.S. Choi, *Obesity and its metabolic complications: the role of adipokines and the relationship between obesity, inflammation, insulin resistance, dyslipidemia and nonalcoholic fatty liver disease.* Int J Mol Sci, 2014. **15**(4): p. 6184-223.
148. Pratley, R.E., *Gene–environment interactions in the pathogenesis of type 2 diabetes mellitus: lessons learned from the Pima Indians.* Proceedings of the Nutrition Society, 1998. **57**(02): p. 175-181.
149. Hales, C.N. and D.J.P. Barker, *Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis*,†.* International Journal of Epidemiology, 2013. **42**(5): p. 1215-1222.
150. Vaxillare, M.F., P., *The Genetics of Type 2 Diabetes: From Candidate Gene Biology to Genome-wide Studies*, in *Textbook of Diabetes*, C.C.S. Holt R.I.G., Flyvbjerg A., Goldstein B.J, Editor. 2010, Balckwell Publishing Ltd. p. 191-214.
151. Stahl, E.A., et al., *Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis.* Nat Genet, 2012. **44**(5): p. 483-9.
152. Wei, W.H., G. Hemani, and C.S. Haley, *Detecting epistasis in human complex traits.* Nat Rev Genet, 2014. **15**(11): p. 722-33.
153. Trerotola, M., et al., *Epigenetic inheritance and the missing heritability.* Hum Genomics, 2015. **9**: p. 17.
154. Ali, O., *Genetics of type 2 diabetes.* World J Diabetes, 2013. **4**(4): p. 114-23.
155. Bevan, S. and H.S. Markus, *Genetics of common polygenic ischaemic stroke: current understanding and future challenges.* Stroke Res Treat, 2011. **2011**: p. 179061.
156. Padmanabhan, J.L., et al., *Polygenic risk for type 2 diabetes mellitus among individuals with psychosis and their relatives.* J Psychiatr Res, 2016. **77**: p. 52-8.
157. Vassy, J.L., et al., *Polygenic type 2 diabetes prediction at the limit of common variant detection.* Diabetes, 2014. **63**(6): p. 2172-82.
158. Vaxillaire, M., 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**(8): p. 1601-10.
159. Wessel, J., et al., *Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility.* Nat Commun, 2015. **6**.
160. Asimit, J. and E. Zeggini, *Rare variant association analysis methods for complex traits.* Annu Rev Genet, 2010. **44**: p. 293-308.
161. Steinthorsdottir, V., et al., *Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes.* Nat Genet, 2014. **46**(3): p. 294-8.

162. Huyghe, J.R., et al., *Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion.* Nat Genet, 2013. **45**(2): p. 197-201.
163. Auer, P.L. and G. Lettre, *Rare variant association studies: considerations, challenges and opportunities.* Genome Med, 2015. **7**(1): p. 16.
164. Igartua, C., et al., *Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma.* Nat Commun, 2015. **6**.
165. Consortium, S.T.D., et al., *Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population.* JAMA, 2014. **311**(22): p. 2305-14.
166. Yang, J.K., et al., *Interactions among related genes of renin-angiotensin system associated with type 2 diabetes.* Diabetes Care, 2010. **33**(10): p. 2271-3.
167. Zuk, O., et al., *The mystery of missing heritability: Genetic interactions create phantom heritability.* Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1193-8.
168. Hsieh, C.H., et al., *Analysis of epistasis for diabetic nephropathy among type 2 diabetic patients.* Hum Mol Genet, 2006. **15**(18): p. 2701-8.
169. Zhou, J.B., et al., *Interaction of Wnt pathway related variants with type 2 diabetes in a Chinese Han population.* PeerJ, 2015. **3**: p. e1304.
170. Wiltshire, S., et al., *Epistasis between type 2 diabetes susceptibility Loci on chromosomes 1q21-25 and 10q23-26 in northern Europeans.* Ann Hum Genet, 2006. **70**(Pt 6): p. 726-37.
171. Cordell, H.J., et al., *Multilocus linkage tests based on affected relative pairs.* Am J Hum Genet, 2000. **66**(4): p. 1273-86.
172. Cox, N.J., et al., *Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans.* Nat Genet, 1999. **21**(2): p. 213-5.
173. Frost, H.R., C.I. Amos, and J.H. Moore, *A global test for gene-gene interactions based on random matrix theory.* Genet Epidemiol, 2016.
174. Colak, R., et al., *JBASE: Joint Bayesian Analysis of Subphenotypes and Epistasis.* Bioinformatics, 2016. **32**(2): p. 203-10.
175. Leontovich, A.A., R.V. Intine, and M.P. Sarras, Jr., *Epigenetic Studies Point to DNA Replication/Repair Genes as a Basis for the Heritable Nature of Long Term Complications in Diabetes.* J Diabetes Res, 2016. **2016**: p. 2860780.
176. Dayeh, T.A., et al., *Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets.* Diabetologia, 2013. **56**(5): p. 1036-46.
177. Dayeh, T., et al., *Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion.* PLoS Genet, 2014. **10**(3): p. e1004160.
178. Volkov, P., et al., *A Genome-Wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits.* PLoS One, 2016. **11**(6): p. e0157776.
179. Jacobsen, S.C., et al., *Effects of short-term high-fat overfeeding on genome-wide DNA methylation in the skeletal muscle of healthy young men.* Diabetologia, 2012. **55**(12): p. 3341-9.
180. Nitert, M.D., et al., *Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes.* Diabetes, 2012. **61**(12): p. 3322-32.
181. Ronn, T., et al., *A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue.* PLoS Genet, 2013. **9**(6): p. e1003572.
182. Santos, J.M., S. Tewari, and S.A. Benite-Ribeiro, *The effect of exercise on epigenetic modifications of PGC1: The impact on type 2 diabetes.* Med Hypotheses, 2014. **82**(6): p. 748-53.
183. Chambers, J.C., et al., *Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study.* Lancet Diabetes Endocrinol, 2015. **3**(7): p. 526-34.
184. Liu, L., Y. Li, and T.O. Tollefsbol, *Gene-environment interactions and epigenetic basis of human diseases.* Curr Issues Mol Biol, 2008. **10**(1-2): p. 25-36.
185. Smith, C.J. and K.K. Ryckman, *Epigenetic and developmental influences on the risk of obesity, diabetes, and metabolic syndrome.* Diabetes Metab Syndr Obes, 2015. **8**: p. 295-302.
186. Brasacchio, D., et al., *Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase enzymes associated with gene-activating epigenetic marks that coexist on the lysine tail.* Diabetes, 2009. **58**(5): p. 1229-36.
187. Fernandez-Valverde, S.L., R.J. Taft, and J.S. Mattick, *MicroRNAs in beta-cell biology, insulin resistance, diabetes and its complications.* Diabetes, 2011. **60**(7): p. 1825-31.
188. Davidson, B.L. and P.B. McCray, Jr., *Current prospects for RNA interference-based therapies.* Nat Rev Genet, 2011. **12**(5): p. 329-40.

189. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
190. Delgado-Rodriguez, M. and J. Llorca, *Bias.* J Epidemiol Community Health, 2004. **58**(8): p. 635-41.
191. Freedman, L.S., et al., *Dealing with dietary measurement error in nutritional cohort studies.* J Natl Cancer Inst, 2011. **103**(14): p. 1086-92.
192. Kelly, P., et al., *Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study.* Int J Behav Nutr Phys Act, 2011. **8**: p. 44.
193. Ewald, B., M. McEvoy, and J. Attia, *Pedometer counts superior to physical activity scale for identifying health markers in older adults.* Br J Sports Med, 2010. **44**(10): p. 756-61.
194. Moffitt, T.E., A. Caspi, and M. Rutter, *Strategy for investigating interactions between measured genes and measured environments.* Arch Gen Psychiatry, 2005. **62**(5): p. 473-81.

CHAPTER 8: APPENDICES

8.1 Overview of the Malaysian Cohort in "Cohort Profile: The Malaysian Cohort (TMC) Project: a Prospective Study of Non-communicable Diseases in a Multi-ethnic Population

Int. J. Epidemiol. Advance Access published April 11, 2014

International Journal of Epidemiology, 2014, 1–9
doi: 10.1093/ije/dyu089
Cohort profile

Cohort profile

Cohort profile: The Malaysian Cohort (TMC) project: a prospective study of non-communicable diseases in a multi-ethnic population

Rahman Jamal,[1]* Syed Zulkifli Syed Zakaria,[2] Mohd Arman Kamaruddin,[1] Nazihah Abd Jalal,[1] Norliza Ismail,[1] Norkhamiwati Mohd Kamil,[1] Noraidatulakma Abdullah,[1] Norhafizah Baharudin,[1] Noor Hamidah Hussin,[3] Hanita Othman,[3] Nor Muhammad Mahadi[4] and the Malaysian Cohort Study Group

[1]UKM Medical Molecular Biology Institute (UMBI), [2]Department of Paediatrics, Faculty of Medicine, [3]Department of Pathology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Cheras, 56000 Kuala Lumpur, Malaysia and [4]Malaysia Genome Institute, Bangi, Selangor, Malaysia

*Corresponding author. UKM Medical Molecular Biology Institute, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Cheras, 56000 Kuala Lumpur, Malaysia. E-mail: rahmanj@ppukm.ukm.edu.my

Accepted 19 March 2014

Abstract

The Malaysian Cohort study was initiated in 2005 by the Malaysian government. The top-down approach to this population-based cohort study ensured the allocation of sufficient funding for the project which aimed to recruit 100 000 individuals aged 35–70 years. Participants were recruited from rural and urban areas as well as from various socioeconomic groups. The main objectives of the study were to identify risk factors, to study gene-environment interaction and to discover biomarkers for the early detection of cancers and other diseases. At recruitment, a questionnaire-based interview was conducted, biophysical measurements were performed and biospecimens were collected, processed and stored. Baseline investigations included fasting blood sugar, fasting lipid profile, renal profile and full blood count. From April 2006 to the end of September 2012 we recruited a total of 106 527participants. The baseline prevalence data showed 16.6% participants with diabetes, 46.5% with hypertension, 44.9% with hypercholesterolaemia and 17.7% with obesity. The follow-up phase commenced in June 2013. This is the most comprehensive and biggest cohort study in Malaysia, and has become a valuable resource for epidemiological and biological research. For information on collaboration and also data access, investigators can contact the project leader at (rahmanj@ppukm.ukm.edu.my).

Key Messages

- This multi-ethnic cohort has provided comparative prevalence rates among the major ethnic groups in Malaysia.
- The prevalence data confirmed the increasing trends of type 2 diabetes, hypertension and hypercholesterolaemia in Malaysia.
- The comparison of the urban and rural populations showed similarity in terms of prevalence of lifestyle diseases due to modernization

Why was the cohort set up?

Malaysia's population of 28.3 million, based on the 2010 national census, is multi-ethnic with the three major ethnic groups making up 95% of the total population.[1] Malays contribute to 63.1% of the population, Chinese 24.6% and Indians 7.3% and the rest is made up of other smaller ethnic groups in East and West Malaysia plus a small population of aborigines.

Non-communicable diseases are fast emerging and becoming the major cause of morbidity and deaths in Malaysia, similar to that in the USA or other developed nations. It is clear that with the increasing modernization and standard of living in Malaysia since its independence in 1957, there has been a major change in lifestyle which includes diet as well as physical activity. The data from the National Health Morbidity Survey II (NHMS II) in 1996 conducted by the Ministry of Health Malaysia showed an 8.3% prevalence of diabetes among the adult population aged $\geq$18 years.[2] The NHMS III in 2006 showed that the prevalence of diabetes has increased to 11.6% and this increased further in the NHMS IV to 15.2%.[3,4] This is rather alarming and there is a similar pattern of increasing diabetes prevalence elsewhere in Asia.[5] For hypertension, the prevalence in the NHMS II was 33% and this increased to 42.6% in the NHMS III. The trend is again similar to some of our Asian neighbours like Thailand and Singapore.[6–8]

As part of the government's increasing efforts to address and investigate the rising trends of non-communicable diseases, the cabinet approved The Malaysian Cohort study in 2005. The top-down approach ensured funding was given to sustain the project at least for the first 5 years. The study proposal was prepared by a team of local experts from various disciplines. Malaysia is a member of The Asia Cohort Consortium whose membership includes South Korea, Japan, China, Taiwan, Singapore, India and the USA.

The Malaysian Cohort aimed to recruit a total of 100 000 individuals from the various ethnic groups. This number is smaller in proportion to the population when compared with the UK Biobank study which has recruited 500 000 participants from a population of 50 million. Nevertheless, we believe it has become a valuable cohort to have, that is now a national resource for researchers in Malaysia as well as providing us with an opportunity to collaborate with international institutions. We have completed the recruitment of the targeted number of participants and we would like to report and share our experience and baseline data with others. As one of the newest cohorts amongst developing nations, we also would like to share the unique experience and the challenges in developing such a study in a tropical and multi-ethnic country like Malaysia.

The primary objectives of TMC project are: (i) to study and determine the roles and interaction of genes, environment and lifestyle in various diseases through a large-scale population cohort study; (ii) to discover biomarkers for cancers and other diseases using the genomics and proteomics approach which would eventually lead to early detection and prevention of diseases; (iii) to consolidate and sustain the initiative for research in life sciences through a systematic discovery programme and also international collaborative research; and (iv) to establish a rich database of information and a bank of biospecimens which will become a national resource for research.

Who is in the cohort?

The Malaysian Cohort study was designed to recruit a total of 100 000 participants aged 35–70 years. The study was approved by the institutional review and ethics board of the Universiti Kebangsaan Malaysia. The study approach included using an interview-based questionnaire and various biophysical measurements plus the collection, processing and storage of biospecimens.

Sampling

The cohort sampling was performed using a mixed approach of voluntary participation (through advertisements and publicity campaigns) as well as cluster and targeted sampling. The cluster sampling was used for the rural areas. The rural areas were chosen from the government's Federal Land Development Authority (FELDA) agricultural scheme which was set up in 1956 and focused on the farming of rubber and oil palm. There are currently about

Table 1. Demographic characteristics of the 106 527 participants in The Malaysian Cohort (2006–12) compared with the general Malaysian population (Census 2010[a])

Demographic	TMC (2006–12)			Malaysian population (Census 2010[a])		
	Place of residence			Place of residence		
	Number of participants	Urban (%)	Rural (%)	Number of people	Urban (%)	Rural (%)
Gender						
Male	44 897	71.8	28.2	14 562 638	70.7	29.3
Female	61 630	71.1	28.9	13 771 497	71.4	28.6
Ethnicity						
Malay	46 782	52.4	47.6	14 191 720	66.6	33.4
Chinese	34 624	96.8	3.2	6 392 636	91.0	9.0
Indian	16 218	86.8	13.2	1 907 827	89.1	10.9
Other	8903	45.0	55.0	5 841 952	54.0	46.0
Age range (years)						
35–44	30 293	80.0	20.0	3 690 093	74.0	26.0
45–54	45 909	70.6	29.4	2 974 602	71.6	28.4
55–64	29 074	64.1	35.9	1 888 618	68.6	31.4
65–70	1251	65.3	34.7	1 427 340	64.4	35.6

[a]see *Population Distribution and Basic Demographic Characteristics 2010* (Department of Statistics Malaysia, 2010[1]).

112 000 settlers working in 103 of these settlements throughout Malaysia, and a total of 75 settlements were sampled. A total of 25 907 invitations were sent out to those who fulfilled the age criteria and 19 467 people (75.1%) responded and were recruited. The FELDA cohort is a relatively non-mobile population and provided an advantage for future follow-up and visits. For the urban areas, the participants were recruited from publicity events which were held in cities, towns, government offices, private agencies and housing areas as well as newspaper advertisements. Between April 2006 and September 2012, a total of 106 527 participants were enrolled into the study. The demographic characteristics of the participants and the comparison with the Malaysian population (as of Census 2010) are shown in Table 1.

The inclusion criteria included being a Malaysian citizen and in possession of a valid identification card, not suffering from any acute illness at the time of study and giving informed consent to the study. Those excluded include those with debilitating illnesses including cancers and those who refused consent. A four-layered written informed consent was taken which covers consent for: (i) the study interview; (ii) the biophysical examination; (iii) blood taking, baseline blood tests and storage of biospecimens; and (iv) future research.

Recruitment centres

The main recruitment centre was based at The Malaysian Cohort office at the Universiti Kebangsaan Malaysia Medical Centre (UKMMC) in Kuala Lumpur. We also had two mobile teams recruiting in the other cities, towns, housing areas and the rural areas. Each of the three recruitment teams consisted of 20–24 personnel including enumerators, phlebotomists, laboratory technicians and data assistants. The mobile teams were also equipped with a mobile laboratory to ensure the preservation of biospecimens in rural areas where electricity supply was a problem, and transportation vans to transport the biospecimens within 24 h from the recruitment sites to the central processing site at the UKMMC. For recruitment in East Malaysia, biospecimens were transported via air shipment.

How are the participants being followed up?

Follow-up and endpoints

Each participant was given a health diary to fill up and return to the TMC office every 6 months. This was to record all illnesses, visits to health facilities, medications and procedures, cost of each treatment and source of payment. Due to the low percentage of the return of these self-report forms, we decided to set up a team to call each participant every 6 months and interviewed them based on the health diary. All participants have either a home phone or a mobile phone. This approach has been successful in getting the follow-up data by phone in 70% of the participants. We have not managed to contact by phone a total of 31 957 (30%) participants, and the reasons for this include not answering the phone (43.2%), voice mail response (21.6%), no ringing tone (15.5%), number not in service

Table 2. Socio-demographic and health differences between those successfully and unsuccessfully followed up by telephone among 106 527 participants of The Malaysian Cohort[a]

Demographic and health differences	Successful (N=74 653)			Unsuccessful (N=31 874)			P-value
	Number of participants	Urban (%)	Rural (%)	Number of participants	Urban (%)	Rural (%)	
Gender							
Male	32 255	74.9	25.1	12 642	64.0	36.0	$\chi^2 = 115.1$
Female	42 398	74.4	25.6	19 232	63.8	36.2	$P < 0.001$
Ethnicity							
Malay	31 603	55.4	44.6	15 179	46.2	53.8	
Chinese	25 764	96.9	3.1	8860	96.2	3.8	$\chi^2 = 1093.9$
Indian	12 026	87.9	12.1	4192	83.5	16.5	$P < 0.001$
Other	5260	51.0	49.0	3643	36.4	63.6	
Risk factors							
Hypertension	34 464	70.7	29.3	14 954	58.8	41.2	
Diabetes mellitus	12 072	68.4	31.6	5523	57.9	42.1	$\chi^2 = 221.5$
High cholesterol	33 106	72.6	27.4	14 014	62.3	37.7	$P < 0.001$
Obesity	12 946	67.9	32.1	5820	57.5	42.5	

[a]All data are row percentages.

(11.8%), wrong number (5.4%) and missing contact number (2.5%). Migration or change of address could also be a cause. The differences between the group which was successfully followed up and the group which we failed to contact are shown in Table 2. A comprehensive follow-up is targeted every 5 years, where each participant will again interviewed, biophysical measurements repeated and biospecimens collected. The invitations to the 5-yearly revisits are issued by phone and through invitation letters posted to their addresses. For the 30% non-responders, our mobile teams will trace them via home visits.

In Malaysia, every citizen is provided with a national identification card (IC) which has a unique number for each individual. For the mortality data, the IC numbers of the participants were sent every 6 months to the National Registration Department (NRD). The NRD provided us with the mortality data and the cause of death.

What has been measured?

Questionnaire and interview

The questionnaire was developed by The Malaysian Cohort Study Group with the assistance of advisers from the Asia Cohort Consortium. Several questionnaires from the Korean Cohort study, the Singapore Chinese Health Study and the Fred Hutchinson Cancer Research Centre, USA, were used with permission as references. The questionnaire covered information on demography, occupational history, use of tobacco and alcohol, diet and physical activity, menstrual and reproductive history (women) and medical history. The diet component consisted of a 24-h recall and a 2-day food record. The physical activity questionnaire was adapted from the short version of the international physical activity questionnaire (IPAQ). The questionnaire was uploaded onto tablet personal computers with touch-screen features. Key pop-up features included a data dictionary as well as a digital diet album, to assist both the enumerators and the participants. Each participant was interviewed face to face at the central recruitment centre or at the mobile sites by a trained interviewer.

Quality control of interviews

We introduced a quality control system for the data obtained from the interviews. Each interview was recorded, with consent, using the tablet computer recording system as well as an MP3 player. Every interview recording was listened to and audited by an independent enumerator. The errors were coded and rectified accordingly.

Biophysical measurements

Each participant had the following measurements taken: height using the Harpenden stadiometer, weight (Seca weighing scale), BMI, waist and hip circumference, waist-to-hip ratio, body composition analysis using the InBody 720 system (Biospace), lung function test using a spirometer (model SP260 by Schiller), blood pressure (HEM-907 model by OMRON) and electrocardiogram. Each measurement was taken three times where possible and

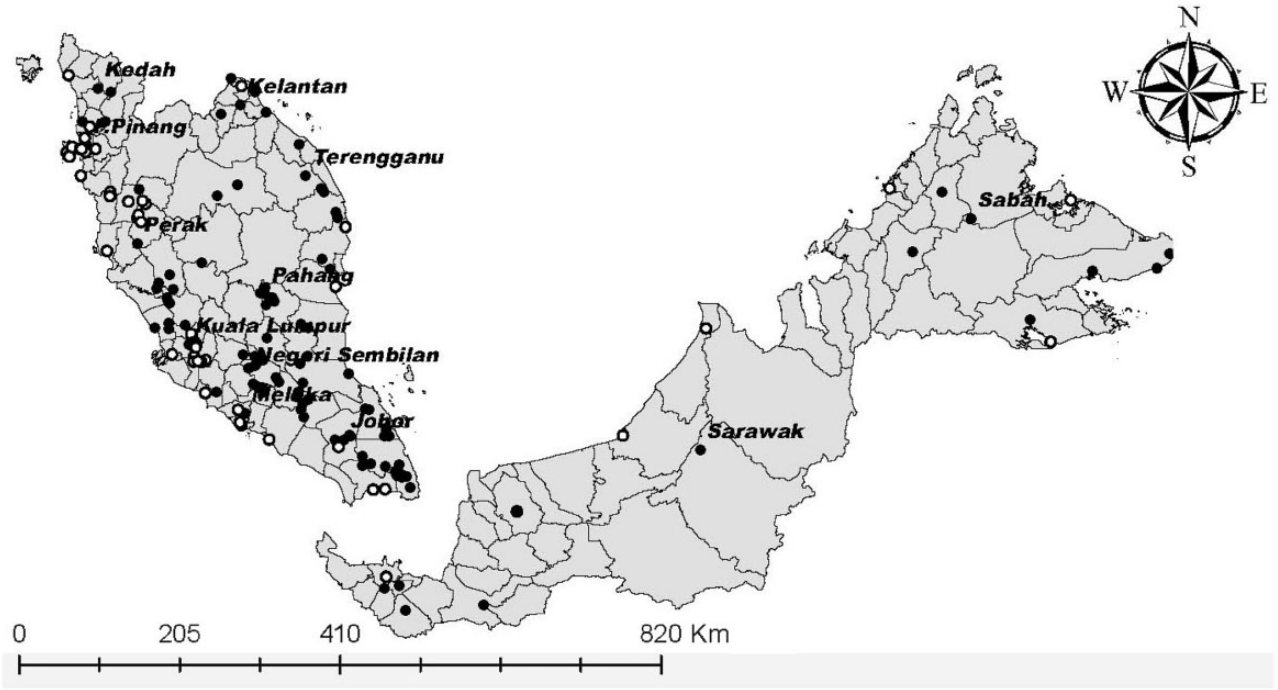


Figure 1. The Malaysian Cohort's 151 recruitment locations, comprising 95 rural (filled circles) and 56 urban (unfilled circles) locations.

underwent quality control processes before the data were uploaded into the database.

Biospecimens

We used the UK Biobank protocol as our reference standards for the collection of blood and urine.[9] All participants came fasted. A total of 40 ml blood and 20 ml urine were donated by each participant. After processing, each participant had 54 cryovials of biospecimens. Half of the cryovials were kept in −80°C freezers and the other half in liquid nitrogen tanks. We used 1-ml Nunc Cryotubes with a 2-D barcoding system (Thermo Scientific, USA) to allow for a systematic inventory system as well as for easy sample tracking and retrieval. The freezers and liquid nitrogen tanks were housed in The Malaysian Cohort Biobank which has an online temperature monitoring system.

Baseline measurements

We measured fasting blood glucose, fasting cholesterol (later full lipid profile after the 50 000th participant), full blood count and renal profile (after the 50 000th participant). Calibration of equipments (Roche Integra and Beckman Coulter) was performed regularly and correlation studies were performed with the hospital's chemical pathology diagnostic laboratory to ensure validity of the results.

Feedback of baseline results to the participants

The results from the biophysical measurements, fasting blood sugar, fasting cholesterol and full blood count were compiled into a one-page summary report and posted to each participant within 2 weeks of recruitment. The report also contained basic explanation on what the normal values were. Those with abnormal results were advised to see their doctor for further investigation and treatment.

Follow-up beginning in June 2013

For the follow-up, we are using the same questionnaire with some minor modifications. Biophysical measurements remain the same but we have added the measurement of cardio-ankle vascular index (CAVI). We are collecting a total of 30 ml blood and 20 ml urine at follow-up. For the blood tests, we added measurement of T4 and thyroxine stimulating hormone (TSH) levels as well as HbA1c levels for those with diabetes.

What has The Malaysian Cohort study found?

Baseline demographic characteristics and habits

As of 30 September 2012, we have recruited a total of 106 527 participants from all over Malaysia. Figure 1 shows

Table 3. Demographic characteristics, educational level, smoking habit, alcohol use and prevalences of diseases according to ethnicity and age group among the 106 527 participants in The Malaysian Cohort (2006–12)[a]

Baseline characteristics ($N = 106\,527$)	Men, by age (years)[a] $N = 44\,897$				Women, by age (years)[a] $N = 61\,630$				Total (%)	Chi-square	
	35–44	45–54	55–64	65–70	35–44	45–54	55–64	65–70		$\chi 2$	*P*-value
Place of residence											
Urban	81.8	73.8	62.0	59.8	78.9	68.4	66.1	73.4	71.4	6.8	<0.001
Rural	18.2	26.2	38.0	40.2	21.1	31.6	33.9	26.6	28.6		
Ethnicity											
Malay	40.2	44.9	50.0	47.6	38.2	46.3	42.7	33.7	43.9		
Chinese	31.4	29.2	28.6	36.3	34.5	32.6	37.6	54.4	32.5	292.4	<0.001
Indian	17.8	17.0	14.6	10.1	16.5	13.6	13.5	7.7	15.2		
Others	10.6	8.9	6.8	6.0	10.8	7.5	6.2	4.2	8.4		
Highest educational level											
University/college	39.5	27.0	20.6	10.3	31.0	20.1	16.9	4.5	24.8		
Secondary school	50.9	51.2	37.0	27.8	56.4	46.9	28.8	21.4	45.6	1173.9	<0.001
Primary school	9.2	20.8	40.3	59.0	11.5	29.3	45.4	53.5	26.6		
No schooling	0.4	1.0	2.1	2.9	1.1	3.7	8.9	20.6	3.0		
Tobacco smoking											
Yes	59.6	58.4	56.6	55.1	5.2	3.2	2.9	4.5	26.6	39 178.0	<0.001
Alcohol drinking											
Yes	11.8	11.4	9.8	9.5	1.8	1.2	1.0	0.8	5.4	4731.5	<0.001
Prevalence											
Hypertension	32.3	46.7	62.1	70.9	25.5	47.4	65.0	80.1	46.5	100.4	<0.001
Malay	30.7	45.3	60.5	72.0	29.6	52.1	69.2	83.6	49.2		
Chinese	31.3	44.6	62.7	68.9	18.6	38.8	58.8	77.2	41.7	608.8	<0.001
Indian	34.1	50.1	63.2	63.5	23.4	45.9	65.3	84.6	45.7		
Others	38.7	54.2	69.9	86.4	35.9	58.7	73.8	81.0	52.9		
Diabetes mellitus	9.9	18.3	26.7	27.3	7.3	15.3	23.3	27.3	16.6	310.3	<0.001
Malay	9.8	18.6	27.0	29.2	8.4	19.1	29.2	32.7	19.2		
Chinese	5.7	10.9	18.6	22.7	3.3	6.7	12.2	22.2	9.1	3376.7	<0.001
Indian	19.9	33.4	45.6	41.9	13.9	25.2	38.0	48.7	28.3		
Others	5.8	12.5	17.5	16.3	5.8	11.8	18.8	9.5	11.1		
High cholesterol	38.7	47.8	51.3	56.5	25.2	45.8	61.6	64.4	44.9	112.1	<0.001
Malay	44.9	54.0	56.6	60.9	30.4	51.5	66.4	66.7	51.0		
Chinese	34.0	42.3	46.3	54.7	21.7	41.4	59.1	63.4	40.8	1381.0	<0.001
Indian	37.8	47.6	49.5	52.7	22.2	40.7	58.6	68.4	41.6		
Others	31.4	35.0	37.9	39.0	22.2	38.8	50.3	52.4	34.4		
Obesity	16.4	14.6	12.1	7.0	19.8	21.3	19.0	10.8	17.7	650.2	<0.001
Malay	18.8	17.7	15.1	10.9	25.7	29.1	26.8	17.5	22.9		
Chinese	11.4	7.8	6.7	1.9	7.8	7.3	7.4	5.1	7.8	3554.8	<0.001
Indian	20.2	18.3	14.0	9.3	29.6	27.9	26.9	23.1	23.4		
Others	15.7	14.5	9.6	4.4	22.6	22.0	18.1	9.5	18.0		
Number of participants	11 451	18 462	14 240	744	18 842	27 447	14 834	507	106 527		

[a]All data are column percentages.

the distribution of all 151 recruitment locations. The breakdown of the participants in terms of age, sex, ethnicity, location (urban or rural), educational level, smoking and alcohol intake is shown in Table 3. There were more women than men. We oversampled the Indians and Chinese to allow us to have more events in these two ethnic groups for comparison with the Malay ethnic group in future research. The prevalences of smoking and alcohol intake among TMC are 26.6% and 5.4%, respectively (Table 3).

Validation studies

Three validation studies are being performed including urine cotinine levels for smoking history, serum

Table 4. Presence of risk factors among the 106 527 participants in the TMC

Risk factors (hypertension, diabetes, hypercholesterolaemia and obesity)	Number of participants	% of total
No risk factor	26 588	25.0
One risk factor		
Hypertension only	11 083	10.4
Diabetes only	1700	1.6
Hypercholesterolaemia only	13 738	12.9
Obesity only	7143	6.7
Sub-total	33 664	31.6
Two risk factors		
Hypertension + diabetes	2112	2.0
Hypertension + hypercholesterolaemia	10 708	10.1
Hypertension + obesity	7519	7.1
Diabetes + hypercholesterolaemia	1875	1.8
Diabetes + obesity	1106	1.0
Hypercholesterolaemia + obesity	4515	4.2
Sub-total	27 835	26.1
Three risk factors		
Hypertension + diabetes + hypercholesterolaemia	3632	3.4
Diabetes + hypercholesterolaemia + obesity	994	0.9
Hypercholesterolaemia + obesity + hypertension	7776	7.3
Obesity + Hypertension + Diabetes only	2304	2.2
Sub-total	14 706	13.8
Four risk factors		
Hypertension + diabetes + hypercholesterolaemia + obesity	3734	3.5
Total	106 527	100.0

carotenoids for fruit and vegetable intake plus a validation study for physical activity using the Actical accelerometer.

Baseline prevalence data and mean values of measurements

The baseline prevalence data of diabetes, hypertension, hypercholesterolaemia and obesity from the 106 527 participants are also shown in Table 3. We used the level of $\geq$7.0 mmol/l as the cut-off point for diabetes [World Health Organization (WHO) criteria] and the 6.21 mmol/l for hypercholesterolaemia (National Institutes of Health, USA).The prevalence of type 2 diabetes of 16.6% is comparable to the 14.6% prevalence obtained from the National Health Morbidity Survey in 2011, although our cohort involved an older starting age group. There are differences in the prevalence of diabetes and obesity between the Chinese (lower prevalence) vs the Malays and Indians. This has provided key opportunities for genome-wide association studies (GWAS) as well as gene-environment-lifestyle comparison between ethnic groups. A genome-wide association study on type 2 diabetes is currently being done. Table 4 shows the presence of risk factors either singly or in combination. A total of 43.4% of the participants have more than one risk factor. The mean values of baseline measurements and blood tests are shown in Table 5. There are differences between the values among men and women and also between the age groups.

Mortality data from 2007–13

Table 6 shows the mortality data and the causes of death since we started recruitment in 2006, up to June 2013. The cause of deaths according to the cancer types are also shown within Table 6.[10] The calculated crude mortality rate for the cohort is 1284 per 100 000 person-years.

What are the main weaknesses and strengths?

Strengths

i. A top-down project approved at the cabinet level ensured the sustained funding from 2005–13.
ii. The establishment of the first large population-based cohort study in Malaysia has comprehensive assessments of exposure, diet and physical activity, biological specimens (blood and urine) and 6-monthly follow-up data.
iii. Many innovative technologies were used, including e-questionnaire (the questionnaire was downloaded to tablet PCs and used by the enumerators to interview the participants), mapping of each participant's address using the Geographical Information System (GIS) and a mobile laboratory for use in the rural communities. The questionnaires were also translated into English, Mandarin and Tamil.
iv. The use of GIS has given us the opportunity to map and layer the environmental data and to facilitate the study of many aspects of diseases including gene-environment interaction.
v. The development of our own in-house Cohort Information Management System (CIMS) manages many key aspects of the study including registration, questionnaire data, biophysical data, results of blood tests, biobank and follow-up data.
vi. Extensive quality control of data includes listening to audio recording of interviews to detect and correct errors, and checking of biophysical data.
vii. The Cohort biobank follows strictly international standards of biobanking and we also use the UK Biobank procedures as a main reference. Our

Table 5. Mean values of baseline measurements and blood tests according to age group and gender among 106 527 participants in The Malaysian Cohort[a]

Mean	Men ($N = 44\,897$) by age group				Women ($N = 61\,630$) by age group				*P*-value
	35–44	45–44	55–64	65–70	35–44	45–44	55–64	65–70	
Systolic blood pressure (mmHg)	124.8 ± 14.5	128.6 ± 16.9	133.2 ± 19.3	137.1 ± 21.3	117.1 ± 15.9	126.5 ± 19.2	133.0 ± 20.5	140.4 ± 22.3	< 0.001
Diastolic blood pressure (mmHg)	82.9 ± 10.9	84.0 ± 11.4	83.1 ± 11.6	81.0 ± 12.0	79.7 ± 11.6	83.0 ± 12.2	82.4 ± 11.6	81.1 ± 11.7	< 0.001
Fasting blood glucose (mmol/l)	5.8 ± 1.7	6.2 ± 2.2	6.5 ± 2.4	6.3 ± 2.0	5.5 ± 1.5	6.0 ± 2.1	6.3 ± 2.4	6.3 ± 2.1	<0.001
Total cholesterol (mmol/l)	5.7 ± 1.1	5.8 ± 1.2	5.9 ± 1.2	5.9 ± 1.2	5.5 ± 0.9	5.9 ± 1.2	6.1 ± 1.2	6.2 ± 1.2	< 0.001
HDL cholesterol (mmol/l)	1.2 ± 0.3	1.2 ± 0.3	1.3 ± 0.3	1.3 ± 0.4	1.5 ± 0.4	1.5 ± 0.4	1.5 ± 0.4	1.6 ± 0.5	< 0.001
LDL cholesterol (mmol/l)	3.7 ± 1.0	3.7 ± 1.0	3.5 ± 1.1	3.6 ± 1.2	3.3 ± 0.9	3.6 ± 1.0	3.7 ± 1.1	3.7 ± 1.1	<0.001
Triglycerides (mmol/l)	1.8 ± 1.3	1.8 ± 1.2	1.7 ± 1.1	1.6 ± 0.9	1.2 ± 0.8	1.4 ± 0.8	1.5 ± 0.9	1.6 ± 0.7	<0.001
BMI (kg/m^2)	26.2 ± 4.4	26.0 ± 4.1	25.6 ± 4.0	24.7 ± 3.6	26.0 ± 5.2	26.5 ± 4.9	26.2 ± 4.8	25.0 ± 4.1	<0.001

[a]Data are means ± SD.

Table 6. Causes of death and number of cases (based on ICD-10) contributing to the mortality in The Malaysian Cohort from commencement of recruitment until June 2013

Main cause of death	Number of cases (%)
Diseases of the circulatory system	440 (32.2)
Neoplasms	
Lung (49 cases)	
Liver (35 cases)	
Breast (35 cases)	
Colorectal (16 cases)	
Stomach (14 cases)	
Nasopharyngeal (13 cases)	
Lymphoma (13 cases)	
Brain (13 cases)	
Ovarian (11 cases)	
Pancreatic (11)	
Unknown (11)	
Other (45)	266 (19.4)
Certain infectious and parasitic diseases	189 (13.8)
Ageing	128 (9.3)
Diseases of the respiratory system	116 (8.5)
Injury, poisoning and similar	55 (4.0)
Endocrine, nutritional and metabolic diseases	48 (3.5)
Unknown/other	31 (2.3)
Diseases of digestive system	29 (2.1)
External causes of morbidity and mortality	24 (1.8)
Diseases of the genitourinary system	20 (1.5)
Symptoms, signs & abnormal clinical and laboratory findings, not elsewhere classified	16 (1.2)
Diseases of the nervous system	4 (0.3)
Diseases of the skin and subcutaneous tissue	2 (0.1)
Total	1368 (100)

bioanalytical laboratory for testing blood sugar, lipid profile, renal profile and full blood count was given the ISO15189 certification from the Department of Standards Malaysia in November 2011.

Weaknesses

i. The The urban Cohort population was somewhat non-representative as we allowed anyone who fulfilled the criteria and those who signed up during our publicity campaigns to contact our call centre and make an appointment to become a participant. However, for the rural community in the agricultural settlements we might have a more representative sample for the population as cluster sampling was used. The data in Table 1 have clearly shown the similarities and differences between the TMC participants and the general Malaysian population. There will certainly be limitations in terms of representativeness and we shall be cautious when using the TMC data in future studies especially in those looking at non-genetic associations.

ii. The failure to contact about 30% of the participants during the 6-monthly phone call was due to migration, transfer of place of work or change in telephone numbers. Measures are being taken to trace them via letters or electronic mail as well as home visits.

iii. Systematic update of exposure data is only now possible as we have completed the target recruitment of 100 000 participants 5 years after the first baseline recruitment. The health diary interview conducted every 6 months covered mainly changes in health status plus treatment.

iv. We had difficulty in obtaining clinical samples at the time of admission for diagnosis of diseases such as cancers. Having tissue samples would certainly add value to future studies.

Can I get hold of the data? Where can I find out more?

Information on The Malaysian Cohort is available at (www.mycohort.gov.com). Requests for other data or

information can be made to the author via e-mail. We welcome national and international collaborations and proposals can be forwarded to (rahmanj@ppukm.ukm.edu.my) and they will then be discussed at the steering committee.

Funding

The project is funded by a top-down grant from the National Biotechnology Division, Ministry of Science, Technology and Innovation (MOSTI), Malaysia (ER-05-01-02-MEB001). The grant is administered by the Malaysia Genome Institute, MOSTI. Additional funding, including infrastructure and utilities, is provided by the Universiti Kebangsaan Malaysia.

Acknowledgements

We thank all the members of The Malaysian Cohort Study Group, all TMC staff members and research assistants. The voluntary participation of all the participants is greatly appreciated. We also thank all the community leaders and the officers from the local authorities involved who have assisted us in the recruitment processes. We are also grateful to the management of the Universiti Kebangsaan Malaysia Medical Centre for providing the space for the TMC office, recruitment centre and the biobank.

The Malaysian Cohort Study Group includes:

Abd Rashid Abd Rahman, Afifah Awang, Ahmad Taufik Jamil, Amal Nasir Mustafa, Amin Ariff Nuruddin, Amin Ismail, Aminuddin Abd Hamid Karim, Amrizal Muhd Nur, Andri Dauni, Atiya Ab Sallam, Christopher Lee, Deepak Rebentisch, Lim Chin Chye, Fauzi Md Anshar, Ghufran Ridzuan, Hashami Bohari, Hayati Aziz, Hejar Abdul Rahman, Ikram Shah Ismail, Ismail Noor, Jamaiyah Haniff, Jamal Hisham Hashim, Jamalludin Ab Rahman, Khadijah Shamsuddin, Krishna Gopal Rampal, Lokman Hakim Sulaiman, Mirnalini Kandiah, Mohamad Ariff Mohd Hashim, Mohd Hashim Omar, Mohd Nizam Isa, Mohd Rizal Abdul Manaf, Muhammad Amir Kamaluddin, Muhamad Nazeri Osman, Nik Mazlan Mamat, Noor Aini Mohd Yusof, Nor Azmi Kamaruddin, Nor Laili Mahat, Norimah Ab. Karim, Osman Ali, Ng Sien Kiat, Poh Bee Koon, Nurmaya Mustafa, Nurul Ain Md Yusuf, Rohaizat Yon, Roslan Harun, Rushidi Ramly, Rusli Ismail, Safiah Mohd Yusof, Salina Abd. Aziz, Salmaan Hussain Inayat Hussain, Salywana A Kamat, Samsuddin Osman Cassim, Saniah Than, Saperi Sulong, Sazaly Abu Bakar, Seri Suniza Sufian, Shahrul Azman Mohd Noah, Sirajuddin Hashim, Suzana Shahar, Syarifah Zainiyah Syed Yahya, Syed Mohd AlJunid, Tan Yoke Hwa, Tengku Mohd Tengku Sembok, Wan Abdul Manan Wan Muda, Wan Ahmad Faisal Wan Salam, Wan Mohamad Wan Bebakar, Wan Nazaimoon Wan Mohamud, Wan Zurinah Wan Ngah, Yasmin Anum Mohd Yusof, Zafar Ahmed, Zainal Ariffin Omar, Zarihah Zain, Zilfalil Alwi, Zulhabri Othman and Zulkifli Ahmad.

Conflict of interest: None declared.

References

1. Department of Statistics Malaysia. *Population Distribution and Basic Demographic Characteristics* 2010. www.statistics.gov.my (18 March 2014, date last accessed).
2. Ministry of Health Malaysia Institute of Public Health. *National Health and Morbidity Survey II*. Kuala Lumpur: Ministry of Health Malaysia Institute of Public Health, 1996.
3. Ministry of Health Malaysia Institute of Public Health. *National Health and Morbidity Survey III*. Kuala Lumpur: Ministry of Health Malaysia Institute of Public Health, 2006.
4. Ministry of Health Malaysia Institute of Public Health. *National Health and Morbidity Survey IV*. Kuala Lumpur: Ministry of Health Malaysia Institute of Public Health, 2011.
5. Ramachandran A, Ma RC, Snehalata C. Diabetes in Asia. *Lancet* 2010;**375**:408–18.
6. Jitnarin N, Kosulwat V, Rojroongwasinkul N, Boonpraderm A, Haddock CK, Poston WS. Prevalence of overweight and obesity in Thai population: results of the National Thai Food Consumption Survey. *Eat Weight Disord* 2011;**16**:242–49.
7. Aekplakorn W, Sangthong R, Kessomboon P *et al*. National Health Examination Survey IV study group. Changes in prevalence, awareness, treatment and control of hypertension in Thai population, 2004-2009: Thai National Health Examination Surveys III-IV. *J Hypertens* 2012;**30**:1734–42.
8. Odegaard AO, Pereira MA, Koh WP *et al*. BMI, all-cause and cause-specific mortality in Chinese Singaporean men and women: the Singapore Chinese Health Study. *PLoS One* 2010;**5**:e14000.
9. UK Biobank Coordinating Centre. *Protocol for a Large-Scale Prospective Epidemiological Resource*. Protocol No: UKBB-PROT-09-06 (Main Phase). Nottingham, UK: Biobank Coordinating Centre, 2007.
10. Jamal R. The Malaysian Cohort Project: Current status and preliminary results. *Symposium 7 Book of Abstracts, Sixth General Assembly of the Asia Pacific Organisation for Cancer Prevention, 26th April, Pullman Hotel, Kuching, Sarawak, Malaysia*, 2012.