# 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 $4.71 \times 10^{-6}$ to $1.35 \times 10^{-8}$), 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, 5$^{th}$ -9$^{th}$ 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