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
Available online 19 May 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