ORIGINAL ARTICLE

Ethnicity-specific polygenic risk score for predicting coronary heart disease mortality in Malaysian patients with type 2 diabetes

Khairul Anwar Zarkasi, PhD[1,2], Nor Azian Abdul Murad, PhD[1], Noraidatulakma Abdullah, PhD[1], Nurul Ain Mhd Yusuf, MSc[1], Nazihah Abd Jalal, BSc[1], Norliza Ismail, BSc[1], Mohd Arman Kamaruddin, MSc[1], Norfazilah Ahmad, PhD[3], Rahman Jamal, PhD[1]

[1]UKM Medical Molecular Biology Institute, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia, [2]Biochemistry Unit, Preclinical Department, Faculty of Medicine and Defence Health, Universiti Pertahanan Nasional Malaysia, Kuala Lumpur, Malaysia, [3]Department of Public Health Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

ABSTRACT

Introduction: Polygenic risk scores (PRS) are used to aggregate single-nucleotide polymorphisms (SNPs) effects across the genome to estimate an individual's risk of coronary heart disease (CHD). However, most PRS models were developed mainly in Caucasian populations, limiting their applicability to Malaysia's diverse populations. This study aimed to develop a PRS to predict CHD mortality among individuals with type 2 diabetes (T2D) in The Malaysian Cohort (TMC).

Materials and Methods: In Phase 1, 224 CHD-death cases and 625 controls were genotyped using Illumina's Infinium Asian Screening Array, followed by genetic imputation. From the genome-wide association analysis, 56 SNPs were identified at suggestive genome-wide significance ($P < 5 \times 10^{-5}$), with 15 SNPs selected to construct the PRS. In Phase 2, 806 T2D participants without known CHD at baseline were followed until death.

Results: The PRS distinguished cases from controls across all ethnicities, with an area under the receiver operating characteristic curve (AUC) indicating good discriminative ability. At 0.85 cut-off score, the PRS achieved 64.7% sensitivity and 73.1% specificity. Those in the high-PRS category had significantly shorter mean survival than those in the low-risk group: observed among Malays (6.60 vs 8.92 years), Chinese (6.20 vs 8.66 years), and Indians (7.21 vs 8.60 years). After adjusting for covariates, high-PRS individuals had increased CHD mortality risk, with hazard ratios of 3.23 (Malays), 9.59 (Chinese), and 9.65 (Indians) (all $p<0.05$).

Conclusion: This study demonstrates the utility of a population-specific PRS in predicting CHD mortality among Malaysian T2D patients, supporting its potential role in precision medicine and long-term risk stratification.

KEYWORDS:
Coronary heart disease; polygenic risk score; single nucleotide polymorphism; type 2 diabetes mellitus

INTRODUCTION

Coronary heart disease (CHD) remains a leading cause of morbidity and mortality worldwide. According to the World Health Organization, CHD accounted for approximately 9.1 million deaths globally in 2021, representing 13% of total mortality.[1] CHD results from the progressive accumulation of atherosclerotic plaque in the coronary arteries, ultimately impairing myocardial tissue perfusion and oxygenation. Type 2 diabetes mellitus (T2D), particularly when poorly controlled, accelerates atherosclerosis and substantially increases the risk of CHD-related mortality.

In Malaysia, T2D poses a growing public health concern, with a national prevalence of 18.3%, which is significantly higher than the global (10.5%), Southeast Asian (8.7%) and Western Pacific (11.9%) averages.[2-3] T2D is strongly associated with both microvascular and macrovascular complications, including CHD. The high prevalence of T2D suggests that Malaysians face an elevated risk of developing CHD, with CHD-related mortality being among the most severe outcomes. Data from the National Heart Association of Malaysia reported a 12% increase in patients requiring percutaneous coronary intervention between 2017 to 2020,[4] coinciding with a rise in CHD-related deaths from 13.9% to 16.1% during the same period.[5]

While traditional risk factors such as age, male sex, smoking, alcohol use, physical inactivity, obesity, dyslipidemia, hypertension, and family history are well established,[6-8] genetic predisposition has emerged as a significant contributor to disease risk. Genome-wide association studies (GWAS) have identified 202 single nucleotide polymorphisms (SNPs) across 109 genomic loci significantly associated with CHD ($P < 5 \times 10^{-8}$).[9] Notably, several SNPs have shown strong associations with CHD risk specifically in T2D patients, including rs10911021 *(GLUL)* ($P = 2.0 \times 10^{-8}$), rs74617384 *(LPA)* ($P = 3.2 \times 10^{-12}$), and rs10811652 (9p21) ($P = 6.0 \times 10^{-11}$).[10,11]

These findings have facilitated the development of polygenic risk scores (PRS), which quantify an individual's genetic risk by summing the effects of multiple risk alleles across the genome. PRS have been shown to predict CHD risk in several large cohorts, including the Framingham Offspring Study

This article was accepted: 14 July 2026
Corresponding Author: Rahman Jamal
Email: rahmanj@hctm.ukm.edu.my

and the Second Northwick Park Heart Study.[12-13] However, the performance and transferability of PRS are often limited by population-specific genetic architectures. Most existing PRS are derived from cohorts of European ancestry, reducing their applicability in other ethnic groups.1[4] Consequently, the number and selection of SNPs included in PRS models vary widely, from dozens to millions, depending on the population and study design.

Despite the growing body of research on PRS for CHD, few studies have focused on multiethnic Southeast Asians populations such as Malaysia's. Therefore, the present study aimed to develop and evaluate a population-specific PRS for predicting CHD mortality in T2D individuals enrolled in The Malaysian Cohort (TMC) project.

MATERIALS AND METHODS

Ethics statement

This study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the Universiti Kebangsaan Malaysia (UKM) Research Ethics Committee (Reference no.: UKM PPI/111/8/JEP-2020-473). Written informed consent was obtained from all participants during their initial recruitment into TMC.

Study design and study population

This study was conducted in two phases. Phase 1 involved a case-control design comprising 224 individuals with type 2 diabetes mellitus (T2D), who died from coronary heart disease (CHD) (T2D+CHD group) and 625 T2D participants who died from non-CHD deaths (T2D group). All participants were identified from The Malaysian Cohort (TMC) project database, recruited between April 2006 and September 2012 from various regions nationwide.[15] Participants with a baseline fasting plasma glucose (FPG) of > 7.5 mmol/L (or 126 mg/dL) and those with established disease history were classified as having T2D.[16]

Causes of death were determined from death certificates obtained from the National Registration Department of Malaysia. CHD was defined as death due to acute coronary syndrome (ACS), unstable angina, ST-elevation myocardial infarction (STEMI), non-ST-elevation myocardial infarction (NSTEMI), or coronary atherosclerosis. Individuals with other forms of diabetes mellitus (e.g., type 1 diabetes mellitus, gestational diabetes mellitus, and maturity-onset diabetes in the young) and those who died from other cardiovascular diseases (e.g., cerebrovascular disease, peripheral artery disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis, and pulmonary embolism) were excluded.

Whole blood samples from participants were retrieved from the TMC Biobank for DNA extraction and genotyping. Findings from genetic analysis in phase 1 were used to develop a polygenic risk score (PRS) for CHD death in individuals with T2D. In phase 2, a retrospective cohort study was conducted among 806 T2D participants who were free from CHD at baseline. The aim was to evaluate the ability of the PRS to predict future CHD-related mortality.

DNA extraction, genotyping, and quality control

Genomic DNA was extracted and purified from whole blood samples using the QIAsymphony automated system (Qiagen, Hilden, Germany). DNA quantification was performed using the Qubit dsDNA BR Assay Kit (Thermo Fischer Scientific, Massachusetts, USA). All DNA samples were normalised to a concentration of 50 ng/µL before further analyses.

Genotyping was conducted using the Infinium Asian Screening Array (ASA)-24 v1.0 Beadchip (Illumina Inc., California, USA), which includes 653,893 SNPs tailored for the East and Southeast Asian populations, including Chinese, Japanese, Korean, Vietnamese, Mongolian, and Malay ethnicities. Initial genotype calling was performed using GenomeStudio software v2.0.5 (Illumina). Quality control of individual genotypes was based on the following criteria: (i) p10GC ≥ 0.38, (ii) call rate ≥ 95%, and (iii) concordance between clinical and genotypic sex. Among all samples, 88,267 SNPs were monomorphic, and 565,626 SNPs were polymorphic.

Genotype data that passed quality control were exported using the PLINK Input Report Plug-in v2.1.4 (Illumina) and uploaded into the Michigan Imputation Server (University of Michigan, USA) for genotype imputation using the following parameters: (i) Imputation software: Minimac4 v1.7.1, (ii) Reference panel: GenomeAsia Pilot (GAsP), (iii) Assembly: GRCh37/hg19, (iv) R2 filter: Off, (v) Phasing: Eagle v2.4 (phased output), (vi) Population: ASN (Asian), and (vii) Mode: Quality control and imputation.

Further quality control was performed using PLINK software v1.90.[17] SNPs were excluded if they met any of the following criteria: minor allele frequency < 0.01, genotype call rate < 95%, significant deviation from the Hardy-Weinberg equilibrium ($P < 1.0 \times 10^{-5}$), excess heterozygosity (± 8 standard deviations [SD] from the mean), discordance between clinical and genotypic sex, or evidence of cryptic relatedness (identity-by-descent [IBD] proportion > 0.1875, corresponding to mid-point between second and third-degree relatives).

Population stratification was assessed via principal component analysis (PCA), comparing study participants to references samples from the Singaporean Genome Variation Project (SGVP). Individuals who did not cluster within ± 6 SD from the mean of their ancestral group along the first two principal components were excluded.

Clinical data

Clinical data were obtained from the TMC Biobank database and included sociodemographic variables (e.g., sex, ethnicity, marital status, education level, employment status, residential area), lifestyle and family history (i.e., smoking status, alcohol intake, use of antihypertensive medications, known case of CHD and family history of CHD), age at TMC recruitment and age of death. Biophysical and laboratory parameters were also included: systolic (SBP) and diastolic blood pressures (DBP), body mass index (BMI), waist circumference, waist-hip ratio, fasting blood glucose (FBG), glycated hemoglobin (HbA1c), total cholesterol, triglycerides, LDL-C, and HDL-C. Participants with FBG > 7.0 mmol/L

and/or $HbA_{1c} \geq 7.0\%$ were categorised as having uncontrolled T2D.[18]

Statistical analysis
Statistical analyses of clinical data were performed using IBM SPSS Statistics version 26 (IBM Corporations, New York, USA) and R version 4.3.0 software (The R Foundation for Statistical Computing) within the RStudio environment (version 2023.06.1 build 524; Posit Software, Massachusetts, USA). Continuous data were expressed as mean ± SD, and comparisons between the cases and controls were made using independent t-tests. Categorical variables were summarised as frequencies and percentages (n, %) and compared with the chi-square (χ^2) test or Fisher's exact test, as appropriate. Statistical significance was defined as a two-sided p-value < 0.05.

Genomic data analysis was performed using PLINK software. The association between individual SNP and CHD-related mortality was tested using the "--assoc" command, which applies a chi-square test. After adjusting for genomic inflation (λ_{GC}), SNPs with $P < 5.0 \times 10^{-5}$ (suggestive genome-wide significance) and directly genotyped on the ASA beadchip were selected for polygenic risk score (PRS) development. Logistic regression was applied to estimate the β-value (log-odds) for each SNP, which were then used as weights in the PRS calculation. PRS scores were computed using the "--score" command in PLINK.

In Phase 1, the discriminatory performance of the PRS in classifying CHD death among individuals with T2D was tested using the area under the receiver operating characteristic curve (AUC). Calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test. In phase 2, participants were stratified into low- and high-risk groups based on a predetermined PRS cut-off. The predictive utility of the PRS for future CHD-related mortality was assessed using the Kaplan-Meier survival analysis and Cox proportional hazards regression. Covariates were included in the Cox model to control potential confounding factors.

RESULTS

A total of 849 participants (224 cases, 625 controls) were included in Phase 1 of the study after quality control procedures, comprising 224 CHD death cases and 625 non-CHD death controls. Ethnic composition included 528 Malays (133 cases, 395 controls), 105 Chinese (21 cases, 84 controls), and 216 Indians (70 cases, 146 controls). Principal component analysis (PCA) demonstrated clear separation of the three ethnic groups, each clustering closely with their respective SGVP counterparts, indicating adequate population structure control (figure not shown). Initial genotyping using the Infinium ASA beadchip yielded 653,893 SNPs. After imputation and post-imputation quality control, the final dataset comprised 959,119 high-quality SNPs for association analysis.

Genome-wide associations with CHD death in T2D
Quantile-quantile (Q-Q) analysis indicated no significant inflation in test statistics, suggesting minimal population stratification or confounding, with a genomic inflation factor (λGC) of 0.99 (figure not shown). Among the ~1 million SNPs analysed, rs6765857 (located within the ZBTB38 gene) exhibited the strongest association with CHD death among participants with T2D (OR = 2.81, 95% CI = 1.90–4.17, $P = 7.18 \times 10^{-7}$), followed by rs57987238 in KLF3 (OR = 6.59, 95% CI = 2.98–14.60, $P = 7.74 \times 10^{-7}$). The risk allele for both variants was G, with minor allele frequencies (MAFs) of 0.066 for rs6765857 and 0.018 for rs57987238. Both SNPs were located within intronic regions of their respective genes.

These results corresponded to the prominent peaks observed on the Manhattan plot, particularly on chromosomes 3 and 4 (figure not shown). A total of 56 SNPs met suggestive genome-wide significance ($P < 5.0 \times 10^{-5}$), although none reached the conventional genome-wide significance threshold ($P < 5.0 \times 10^{-8}$) (data not shown). These SNPs were mapped to 23 loci, distributed as follows: 43 in intronic regions, seven intergenic, four upstream, one downstream, and one in the 3' untranslated region. MAFs for the 56 SNPs ranged between 0.010 to 0.383.

Discriminative performance of the polygenic risk score (PRS)
Fifteen lead SNPs were selected from the list of 56 suggestive variants to construct a polygenic risk score (PRS), using their logistic regression β-coefficients as weights (Table I). All selected SNPs were directly genotyped from the ASA chip and represented distinct genetic loci. SNPs with a positive β-values were associated with increased risk of CHD death, whereas negative β-values indicated protective effects. The largest positive β-value was observed for rs71505297 in the *RXRA* gene (β = 2.00) suggesting a strong risk-enhancing effect. Conversely, the most negative β-value was identified for rs10751517 in *HMCN2* (β = –0.95), indicating a potential protective role.

Type 2 diabetic participants who died from CHD had significantly higher mean PRS values compared to those who died from non-CHD causes. This trend was observed both in terms of the overall cohort and within each ethnic group (Figure 1). Discrimination analysis using the AUC demonstrated that PRS could significantly distinguish CHD-related deaths from non-CHD deaths among T2D participants across all ethnicities. The AUC was 0.792 for Malays ($p<0.001$), 0.744 for Chinese ($p=0.001$), 0.792 for Indians ($p<0.001$), and 0.762 for the overall sample ($p<0.001$), indicating strong discriminative ability. Calibration of the PRS using the Hosmer-Lemeshow goodness-of-fit test showed good agreement between observed and expected CHD death frequencies within each ethnicity and the overall sample ($p>0.05$). A clear increasing trend in CHD death prevalence was observed across PRS deciles, supporting the PRS's gradational predictive value (figure not shown).[19]

Baseline characteristics of the retrospective cohort study
Based on the AUC analysis, a PRS cut-off score of 0.85 was selected for risk stratification, yielding 64.7% sensitivity and 73.1% specificity. While the moderate sensitivity and specificity suggest some limitations in individual-level prediction, the PRS may still hold value in identifying population-level risk trends and guiding targeted interventions.

Table I: Selected lead single nucleotide polymorphisms for polygenic risk score

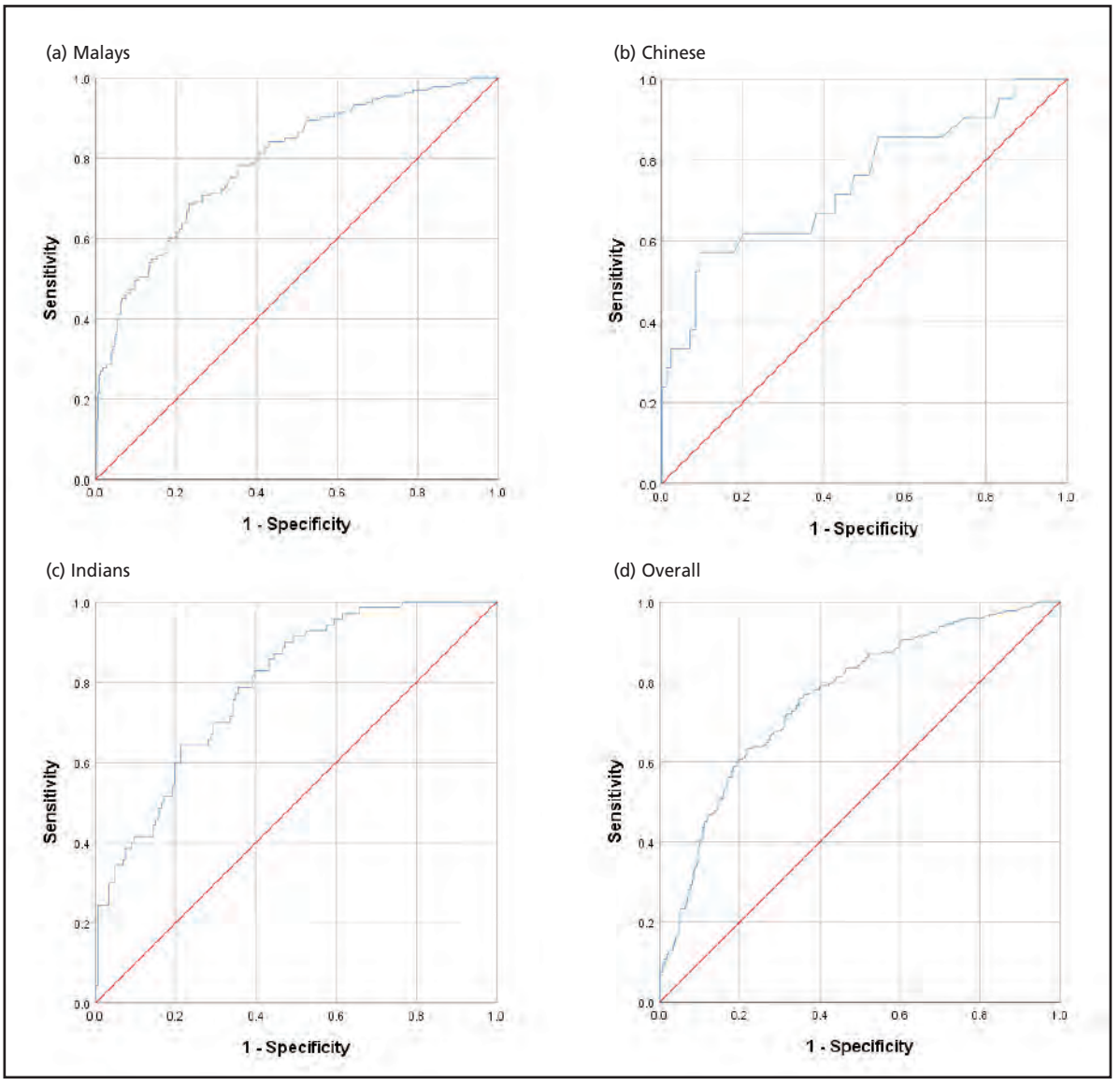
Chr	SNP	A1	A2	Mapped gene	Variant	β
1	rs483202	G	A	GRHL3	Intronic	0.56
1	rs10910660	C	T	PCNX2	Intronic	0.53
1	rs6692683	A	G	EPHB2	Intronic	0.46
2	rs807608	C	T	-	Intergenic	–0.85
3	rs6795197	A	C	ZBTB38	Intronic	0.88
4	rs35887929	C	T	UGT2B10	Intronic	1.24
4	rs10008290	G	A	LOC105377356	Intronic	1.15
5	rs6888803	C	T	LOC105377700	Intronic	–0.75
6	rs34836069	A	G	POLR1C	Intronic	0.52
7	rs2867567	A	G	CALN1	Intronic	–0.48
9	rs10751517	C	T	HMCN2	Intronic	–0.95
9	rs71505297	T	C	RXRA	Intronic	2.00
19	rs56166910	G	A	HAS1	Intronic	–0.49
19	rs1865062	A	G	MLLT1	Intronic	1.49
20	rs6102788	A	G	PTPRT	Intronic	–0.59

Abbreviations: A1 (minor allele), A2 (major allele), Chr (chromosome), SNP (single nucleotide polymorphism).

Table II: The baseline characteristics of the participants with type 2 diabetes mellitus in the retrospective cohort study were based on their polygenic risk categories

Characteristic	High-risk (PRS ≥ 0.85)	Low-risk (PRS < 0.85)	p-value
Age at recruitment (years)	55.22 ± 6.27	56.49 ± 6.01	0.005*
Sex			
Female	94 (29.7%)	223 (70.3%)	0.001*
Male	200 (40.9%)	289 (59.1%)	
Ethnicity			
Malay	117 (23.2%)	387 (76.8%)	< 0.001*
Chinese	11 (10.9%)	90 (89.1%)	
Indian	166 (82.6%)	35 (17.4%)	
Marital status			
Married	266 (36.7%)	458 (63.3%)	0.891
Single	5 (35.7%)	9 (64.3%)	
Divorced/widowed	23 (33.8%)	45 (66.2%)	
Education level			
Primary	153 (33.7%)	301 (66.3%)	0.102
Secondary	128 (41.0%)	184 (59.0%)	
Tertiary	13 (32.5%)	27 (67.5%)	
Working status			
Not working	146 (34.8%)	274 (65.2%)	0.292
Working	148 (38.3%)	238 (61.7%)	
Residency			
Rural	103 (27.1%)	277 (72.9%)	< 0.001*
Urban	191 (44.8%)	235 (55.2%)	
Smoking status			
Not/past smoker	220 (35.4%)	401 (64.6%)	0.257
Active smoker	74 (40.0%)	111 (60.0%)	
Alcohol intake			
No	255 (34.2%)	491 (65.8%)	< 0.001*
Yes	39 (65.0%)	21 (35.0%)	
Diabetes duration			
≤ 5 years	153 (33.3%)	306 (66.7%)	0.033*
> 5 years	141 (40.6%)	206 (59.4%)	
Family history of CHD			
No	90 (30.1%)	209 (69.9%)	0.004*
Yes	204 (40.2%)	303 (59.8%)	
SBP (mmHg)	139.03 ± 21.92	144.69 ± 24.82	0.001*
DBP (mmHg)	80.56 ± 11.49	80.93 ± 12.59	0.674
Body mass index (kg/m²)	26.70 ± 4.43	27.20 ± 4.86	0.151
Waist circumference (cm)	91.98 ± 11.41	91.42 ± 11.26	0.497
Waist-hip ratio	0.94 ± 0.07	0.93 ± 0.07	0.047*
Total cholesterol (mmol/L)	5.79 ± 1.35	6.13 ± 1.51	0.001*
Triglyceride (mmol/L)	2.14 ± 1.33	2.32 ± 2.19	0.147
LDL-C (mmol/L)	3.66 ± 1.21	3.86 ± 1.26	0.025*
HDL-C (mmol/L)	1.15 ± 0.31	1.23 ± 0.38	0.002*
FBG (mmol/L)	10.48 ± 4.42	10.05 ± 4.08	0.169
HbA1c (%)	9.10 ± 2.33	8.98 ± 2.50	0.508

Data are presented as frequency (percentage) and mean ± standard deviation (SD) for categorical and continuous variables. *Significant difference between high and low PRS groups at P < 0.05. Abbreviations: CHD (coronary heart disease), DBP (diastolic blood pressure), FBG (fasting blood glucose), HbA1c (glycated hemoglobin), HDL-C (high-density lipoprotein cholesterol), LDL-C (low-density lipoprotein cholesterol), SBP (systolic blood pressure).

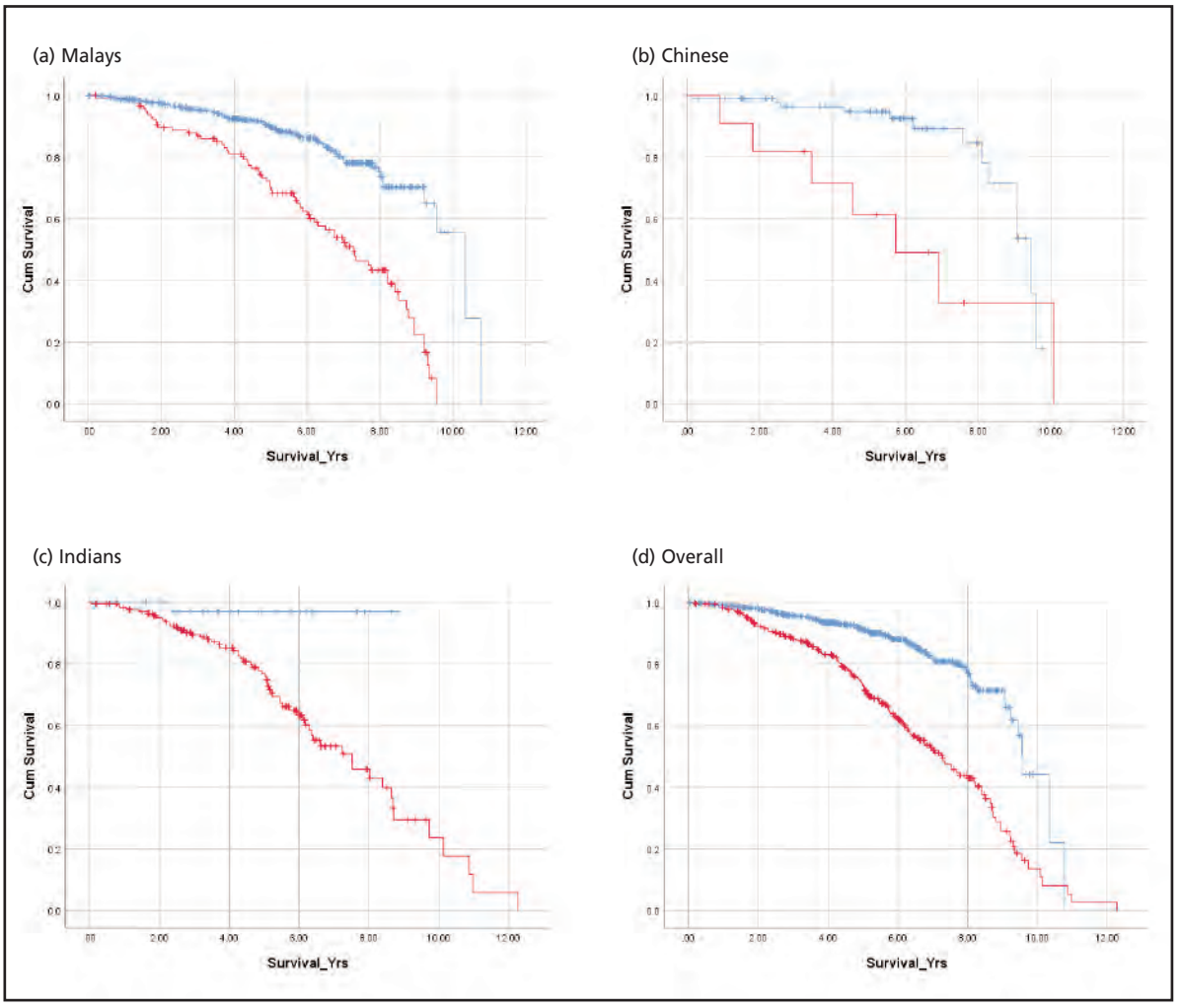


Ethnicity	PRS			p-value
	T2D+CHD	T2D	AUC	
Malay	0.85 ± 0.04	0.82 ± 0.03	0.792	< 0.001*
Chinese	0.84 ± 0.04	0.81 ± 0.02	0.744	0.001*
Indian	0.94 ± 0.05	0.88 ± 0.04	0.792	< 0.001*
Overall	0.88 ± 0.06	0.83 ± 0.04	0.762	< 0.001*

Fig. 1: Discrimination analysis shows the area under the receiver operating characteristic curve and the mean values of polygenic risk score among (a) Malays, (b) Chinese, (c) Indians, and (d) Overall subjects

*Significant results at p<0.05. Legends: Blue line (PRS), red line (reference). Abbreviations: AUC (area under receiver operating characteristic curve), CHD (coronary heart disease), PRS (polygenic risk score), T2D (type 2 diabetes)

To test its clinical utility, a retrospective cohort of 806 T2D participants without prior CHD at baseline was analysed. Participants were stratified into high-risk (PRS≥0.85) and low-risk (PRS<0.85) groups. The high-risk group was significantly younger (55.22 vs 56.49 years, p=0.005), comprised more males (40.9 vs 29.7%, p=0.001), and had a higher proportion of Indian ethnicity (82.6%) compared to Malays (23.2%) and Chinese (10.9%) (p<0.001). Urban residency was also more prevalent in the high-risk group (44.8 vs 27.1%, p<0.001) (Table II). Additionally, the high-risk group had higher prevalence of alcohol consumption (65.0 vs 35.4%, p<0.001), longer diabetes duration (> 5 years: 40.6 vs 33.3%, p=0.033), and greater family history of CHD (40.2 vs 30.1%, p=0.004). Biophysically, the high-risk individuals had lower systolic blood pressure (139.03 vs 144.69 mmHg, p=0.001), higher waist-hip ratio (0.94 vs 0.93, p=0.047), and more favourable lipid profiles, with lower serum total cholesterol (5.79 vs 6.13 mmol/L, p=0.001), LDL-C (3.66 vs 3.86 mmol/L, p=0.025), except for HDL-C (1.15 vs. 1.23 mmol/L, p=0.002).



Ethnicity	Mean survival years (95% CI)		p-value
	High risk PRS	Low risk PRS	
Malays	6.60 (6.08–7.11)	8.92 (8.50–9.35)	< 0.001*
Chinese	6.20 (3.98–8.41)	8.66 (8.15–9.16)	0.011*
Indians	7.21 (6.54–7.88)	8.60 (8.21–8.99)	0.001*
Overall	6.88 (6.47–7.29)	8.88 (8.52–9.25)	< 0.001*

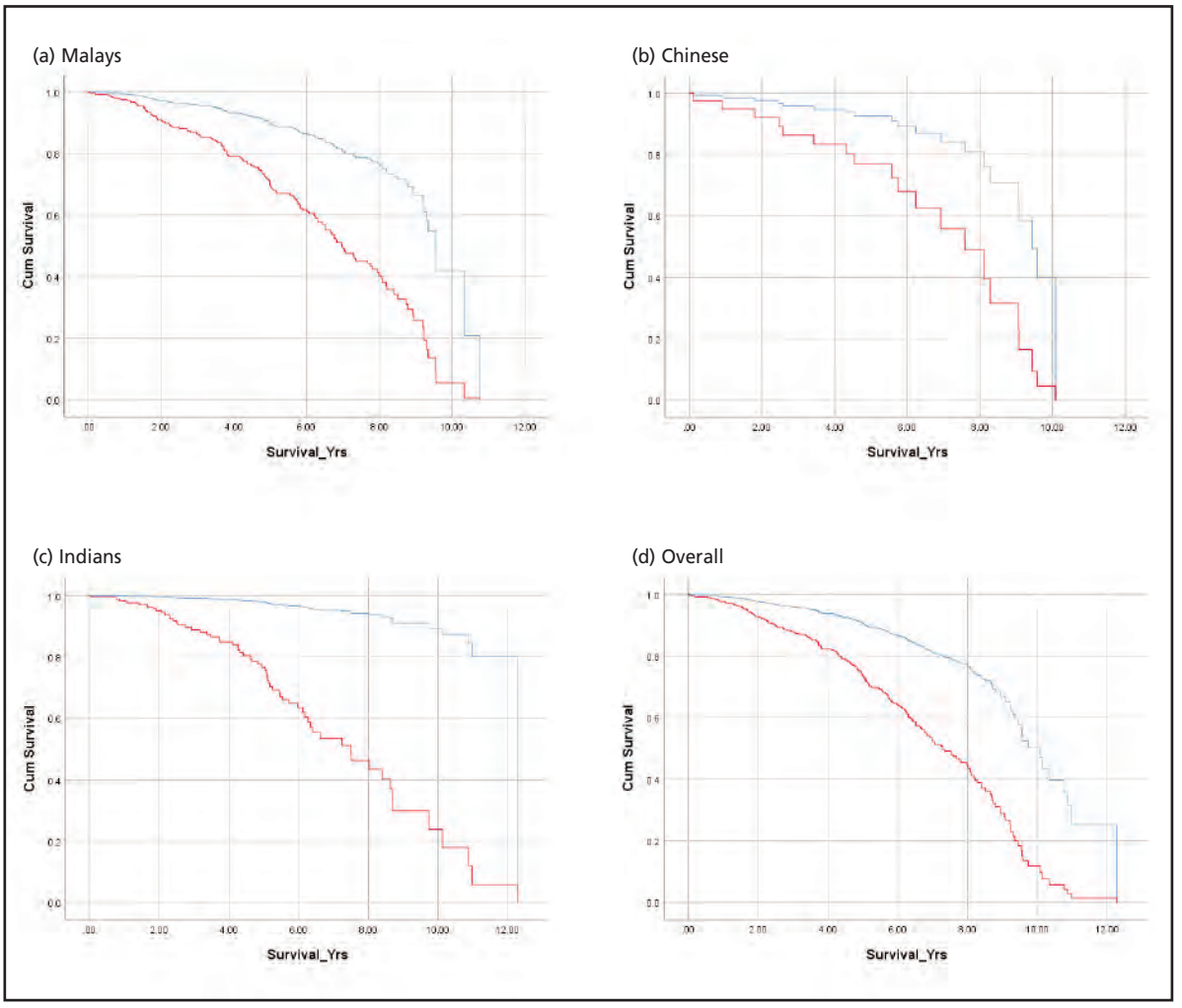
Fig. 2: Kaplan-Meier survival analysis results for the prediction of future CHD death by polygenic risk score among T2D TMC participants in the retrospective cohort study among (a) Malays, (b) Chinese, (c) Indians, and (d) Overall subjects

*Significant results at p<0.05. Abbreviations: CHD (coronary heart disease), CI (confidence interval), Cum Survival (cumulative survival), PRS (polygenic risk score), Survival (survival years). Legends: Blue line (low-risk PRS), red line (high-risk PRS)

Validation of PRS in predicting future CHD death

The predictive utility of the PRS cut-off (≥0.85) for future CHD death among T2D individuals was validated in the retrospective cohort study using Kaplan-Meier (KM) survival analysis and Cox proportional hazards analysis with covariate adjustments. Kaplan-Meier analysis demonstrated significant shorter mean survival durations among high-risk participants across all ethnic groups: Malays (6.60 vs 8.92 years), Chinese (6.20 vs 8.66 years), and Indians (7.21 vs 8.60 years), all with p<0.05 (Figure 2). In the overall samples, high-risk individuals also had significantly reduced survival (6.88 vs 8.88 years, p<0.001).

Cox regression analysis further confirmed that high-risk individuals (PRS ≥ 0.85) had significantly increased hazards of CHD death. After adjusting for relevant covariates (age, sex, ethnicity, urban/rural residence, alcohol consumption, diabetes duration, family history of CHD, systolic blood pressure, waist-hip ratio, total cholesterol, LDL-C, and HDL-C), the adjusted hazard rations were: Malays: HR=3.23 (p<0.001); Chinese: HR=9.59 (p=0.005); Indians: HR=9.65 (p=0.027); Overall sample: HR=3.51 (p<0.001) (Figure 3). These findings underscore the PRS's robust predictive capacity for CHD mortality risk across ethnic groups in the Malaysian T2D population.



Ethnicity	Ref.: Low-risk PRS (< 0.85)					p-value
	β	SE	Wald	df	HR (95% CI)	
Malays	1.17	0.19	36.57	1	3.23 (2.21–4.73)	< 0.001*
Chinese	2.26	0.81	7.80	1	9.59 (1.96–46.90)	0.005*
Indians	2.27	1.02	4.91	1	9.65 (1.30–71.66)	0.027*
Overall	1.26	0.17	54.22	1	3.51 (2.52–4.91)	< 0.001*

Fig. 3: Cox regression analysis results for the prediction of future CHD death by polygenic risk score among T2D TMC participants in the retrospective cohort study among (a) Malays, (b) Chinese, (c) Indians, and (d) Overall subjects

Regression was performed with covariate adjustment (i.e., age at recruitment, gender, ethnicity, residency, alcohol intake, diabetes duration, family history of CHD, SBP, WHR, total cholesterol, LDL-C, and HDL-C). *Significant results at p<0.05. Abbreviations: CHD (coronary heart disease), CI (confidence interval), Cum Survival (cumulative survival), HR (hazard ratio), PRS (polygenic risk score), Survival_Yrs (survival years). Legends: Blue line (low-risk PRS), red line (high-risk PRS).

DISCUSSION

From the 56 SNPs identified with suggestive genome-wide significance in Phase 1 of this study, the strongest association with CHD death among the T2D participants in the TMC was observed for rs6765857, located in the intron of the *ZBTB38* gene (OR=2.81, P = 7.18 × 10^{-7}). This imputed variant has a minor allele frequency of 0.0066. Interestingly, another variant within the same gene that was directly genotyped on the array, rs6795197, also ranked among the top five SNPs most strongly associated with CHD death in our cohort.

The *ZBTB38* gene encodes a zinc finger and BTB domain-containing protein known to regulate gene transcription through binding to methylated CpG dinucleotides.20 Functional studies suggest that *ZBTB38* modulates the expression of approximately 2,000 target genes, and some of these genes have been reported to associate with T2D and CHD, including angiopoietin-like 4 *(ANGPTL4)*, feline sarcoma proto-oncogene *(FES)*, dan serpin family A member 1 *(SERPINA1)*, which are involved in lipid metabolism, vascular function, and inflammation, respectively.[21–25] A pleiotropic GWAS analysis has also identified the 3q23 locus (harboring *ZBTB38*) as a shared genetic susceptibility region for both prostate cancer and T2DM.[22]

Further supporting its relevance, the DHS-MIND study involving 506 Caucasians in The United States, demonstrated that rs6765857 and rs6795197 were significantly associated with serum levels of advanced glycation end-products (AGE) levels,[26] which are known to accumulate in hyperglycaemic conditions and contribute to vascular inflammation and atherogenesis via pathways involving IL-6, IL-8, VCAM1, and ICAM-1.[27–30] These mechanistic insights suggest a plausible biological role for ZBTB38 variants in mediating CHD risk among T2D individuals.

In the context of genome-wide association studies (GWAS), the *ZBTB38* is associated with multiple metabolic traits including body fat percentage,[31] fasting blood glucose,[32] and glycated haemoglobin (HbA1c) levels.[33] Interestingly, despite strong trans-ethnic associations with T2D, genome-wide significance was observed only among Caucasians ($p=1.10 \times 10^{-14}$) and not among Asian populations ($p=3.90 \times 10^{-1}$),[34] underscoring the importance of population-specific genetic studies.

In the current study, a 15-SNP PRS was developed using lead variants with suggestive significance. This PRS demonstrated good discriminative ability (AUC=0.762 for the overall population), with performance consistent across Malay, Chinese and Indian subgroups. Apart from rs6795197 *(ZBTB38)*, other SNPs comprised rs483202 *(GRHL3)*, rs10910660 *(PCNX2)*, rs6692683 *(EPHB2)*, rs807608 (intergenic), rs35887929 *(UGT2B10)*, rs10008290 *(LOC105377356)*, rs6888803 *(LOC105377700)*, rs34836069 *(POLR1C)*, rs2867567 *(CALN1)*, rs10751517 *(HMCN2)*, rs71505297 *(RXRA)*, rs56166910 *(HAS1)*, rs1865062 *(MLLT1)*, and rs6102788 (PTPRT). Many of the SNPs' association with CHD death had not been reported before. However, rs35887929 *(UGT2B10)* has been linked to nicotine metabolism,[35] and rs71505297 *(RXRA)* to levels of azelaic acid levels, a fatty acid with potential anti-atherosclerotic effects,[36-37]

Using a PRS cut-off score of ≥0.85, sensitivity and specificity were 64.7% and 73.1% respectively. While these values indicate moderate diagnostic performance, they remain clinically useful in stratifying patients into high- and low-risk categories for CHD death, particularly in a resource-limited setting where early risk identification can guide preventive strategies. However, the modest sensitivity suggests that a proportion of at-risk individuals may not be identified using this PRS alone, highlighting the need for integration with traditional risk factors in future predictive models. The CARDIoGRAMplusC4D consortium, reported that PRS is usually associated with traditional cardiovascular risk factors, including elevated triglyceride, LDL-C, glucose, and HbA1c, with decreased HDL-C levels.[38]

Importantly, participants in the high-risk PRS group exhibited characteristics not typically associated with higher CHD risk: lower SBP, total cholesterol, and LDL-C levels. This suggests that the PRS may capture genetic risk that operates independently of conventional cardiometabolic markers, possibly through alternative pathways such as inflammation, oxidative stress, or endothelial dysfunction. Despite these differences, locally developed PRS in the present study was able to predict future CHD deaths, similar to prior findings among 401,471 Caucasians in Canada.39 Although using a different set of SNPs, the PRS can predict the incidence of myocardial infarction and all-cause mortality with a remarkable p-value.[39]

It is worth noting that the Cox regression analysis yielded remarkably high hazard ratios for the Chinese (HR=9.59) and Indian (HR = 9.65) subgroups compared to Malays (HR=3.23), accompanied by wide 95% confidence intervals. These observations are direct consequences of smaller subgroup sample sizes and uneven sample distribution within risk categories in Phase 2, such as the small number of Chinese individuals in the high-risk group (n=11) and Indian individuals in the low-risk group (n=35). While the wide confidence intervals imply that the numerical magnitude of these hazard ratios must be interpreted with caution, the persistent statistical significance ($p<0.05$) across all adjusted subgroups underscores a robust and genuine genetic predisposition to accelerated CHD mortality in these populations.

Despite the encouraging findings, this study has several limitations. First, the sample size, although adequate for initial GWAS discovery, was relatively small by global GWAS standards, limiting the detection of SNPs reaching conventional genome-wide significance ($P < 5 \times 10^{-8}$). The selection of lead variants based on a suggestive threshold ($P < 5 \times 10^{-5}$) demands a cautious interpretation regarding individual SNP effects, as it carries an inherently higher risk of type I error. However, this strategy was adopted to capture population-specific genetic variations that would otherwise be ignored. Besides, the risk of false-positive was substantially mitigated by the robust predictive performance of the 15-SNP PRS in the Phase 2 of the study. Second, the use of the Illumina Asian Screening Array, which lacks comprehensive South-Asian genomic content, may have affected variant detection among Indian participants. This limitation was mitigated through imputation using the GenomeAsia Pilot reference panel, which incorporates genetic data from diverse Asian populations.

Nevertheless, this study also has notable strengths. It is the first to identify *ZBTB38* as a top candidate gene associated with CHD death among Malaysian individuals with T2D, and it introduces a population-specific PRS with predictive validity in a multi-ethnic Asian cohort. These findings provide a foundation for the development of precision medicine tools aimed at improving cardiovascular risk stratification and management among T2D patients in Malaysia and potentially other Asian populations.

CONCLUSION

This study identified 56 SNPs with suggestive genome-wide significance for CHD death among T2D participants in the TMC, with rs6765857 in *ZBTB38* emerging as the strongest genetic signal. A 15-SNP PRS demonstrated potential clinical utility in predicting CHD death risk in this population, independent of traditional cardiovascular risk factors. The findings support the integration of PRS into precision medicine approaches for secondary prevention of CHD among individuals with T2D, particularly in multi-ethnic

Southeast Asian populations. Future studies with larger sample sizes and prospective validation are warranted to confirm these associations and enhance PRS performance.

ACKNOWLEDGEMENTS

This research was supported by the UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia (UKM), through grant number PDE48 for The Malaysian Cohort (TMC) project. The publication fee was funded by the Research Management Centre, Universiti Pertahanan Nasional Malaysia (UPNM). The authors extend their sincere gratitude to the postgraduate students, staff, and academic personnel of UMBI, UKM, as well as staff of TMC, particularly Ts. Muhiddin Ishak, Mdm. Zam Zureena Mohd Rani, and Mr. Mohd Ridhwan Abd Razak for valuable assistance and support, both direct and indirect, throughout the course of this research.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. World Health Organization. The Top 10 Causes of Death. [cited April 2025] Accessed from: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
2. National Institute of Health, Shah Alam, Selangor, Malaysia. 2020. National Health and Morbidity Survey 2019: Technical Report – Volume I: NCDs – Non-Communicable Diseases: Risk Factors and other Health Problems.
3. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022; 183: 109119.
4. National Heart Association of Malaysia, Kuala Lumpur, Malaysia. 2023. Annual Report of the NCVD-PCI Registry, Year 2019–2020.
5. Department of Statistics Malaysia. Statistics on Causes of Death, Malaysia, 2023. [cited April 2025] Accessed from: https://www.dosm.gov.my/portal-main/release-content/statistics-on-causes-of-death-malaysia-2023.
6. Dendup T, Feng X, Clingan S, Astell-Burt T. Environmental risk factors for developing type 2 diabetes mellitus: A systematic review. Int J Environ Res Public Health 2018; 15(1): 78.
7. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. J Cell Physiol 2019; 234(10): 16812–23.
8. Ab Zayin AA, Zainol Z, Nohuddin PNE, Mohamed H. Peramalan risiko diabetes menggunakan aplikasi MyDiabeticRisk. Zulfaqar J Def Sci Eng Technol 2023; 6(2): 1-12.
9. McPherson R, Tybjaerg-Hansen A. Genetics of coronary artery disease. Circ Res 2016; 118(4): 564-78.
10. Qi L, Qi Q, Prudente S, Mendonca C, Andreozzi F, di Pietro N, et al. Association between a genetic variant related to glutamic acid metabolism and coronary heart disease in individuals with type 2 diabetes. JAMA 2013; 310(8): 821-8.
11. Fall T, Gustafsson S, Orho-Melander M, Ingelsson E. Genome-wide association study of coronary artery disease among individuals with diabetes: The UK Biobank. Diabetologia 2018; 61(10): 2174-9.
12. Thanassoulis G, Peloso GM, Pencina MJ, Hoffmann U, Fox CS, Cupples LA, et al. A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium. Circ Cardiovasc Genet 2012; 5(1): 113-21.
13. Beaney KE, Cooper JA, Ullah Shahid S, Ahmed W, Qamar R, Drenos F, et al. Clinical utility of a coronary heart disease risk prediction gene score in UK healthy middle aged men and in the Pakistani population. PLoS One 2015; 10(7): 1-13.
14. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat Genet 2021; 53(4): 420-5.
15. Jamal R, Syed Zakaria SZ, Kamaruddin MA, Abd Jalal N, Ismail N, Mohd Kamil N, et al. Cohort profile: The Malaysian Cohort (TMC) project: A prospective study of non-communicable diseases in a multi-ethnic population. Int J Epidemiol 2015; 44(2): 423-31.
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(10): 1377-84.
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 2015; 4: 7.
18. Ministry of Health Malaysia, Putrajaya, Malaysia. 2020. Clinical Practice Guidelines: Management of Type 2 Diabetes Mellitus, 6th Edition.
19. Schulz C, Padmanabhan S. Methods to assess genetic risk prediction. In: Touyz RM, Schiffrin EL, Editors. Hypertension: Methods in molecular biology, vol. 1527. Humana Press; 2017: 27-40.
20. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 2016; 54: 1.30.1-33.
21. Ding G, Lu W, Zhang Q, Li K, Zhou H, Wang F, et al. ZBTB38 suppresses prostate cancer cell proliferation and migration via directly promoting DKK1 expression. Cell Death Dis 2021; 12(11): 998.
22. Ray D, Chatterjee N. A powerful method for pleiotropic analysis under composite null hypothesis identifies novel shared loci between type 2 diabetes and prostate cancer. PLoS Genet 2020; 16(12): e1009218.
23. Yang Q, Yin RX, Cao XL, Huang F, Zhou YJ, Chen WX. ANGPTL4 variants and their haplotypes are associated with serum lipid levels, the risk of coronary artery disease and ischemic stroke and atorvastatin cholesterol-lowering responses. Nutr Metab 2018; 15(1): 70.
24. Yang X, Yang W, McVey DG, Zhao G, Hu J, Poston RN, et al. FURIN expression in vascular endothelial cells is modulated by a coronary artery disease-associated genetic variant and influences monocyte transendothelial migration. J Am Heart Assoc 2020; 9(4): e014333.
25. Liu G, Liao W, Lv X, Huang L, He M, Li L. A potential coagulation-related diagnostic model associated with immune infiltration for acute myocardial infarction. Genes Immun 2024; 25(6): 471–82.
26. Adams JN, Raffield LM, Martelle SE, Freedman BI, Langefeld CD, Carr JJ, et al. Genetic analysis of advanced glycation end products in the DHS MIND study. Gene 2016; 584(2): 173-9.
27. Hashim Z, Zarina S. Advanced glycation end products in diabetic and non-diabetic human subjects suffering from cataract. AGE 2011; 33(3): 377-84.
28. Kilhovd BK, Berg TJ, Birkeland KI, Thorsby P, Hanssen KF. Serum levels of advanced glycation end products are increased in patients with type 2 diabetes and coronary heart disease. Diabetes Care 1999; 22(9): 1543-8.
29. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000; 28(1): 27-30.
30. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. Nature 2011; 473(7347): 317-25.
31. Harris BHL, Di Giovannantonio M, Zhang P, Harris DA, Lord SR, Allen NE, et al. New role of fat-free mass in cancer risk linked with genetic predisposition. Sci Rep 2024; 14(1): 7270.

32. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. Nat Genet 2021; 53(6): 840-60.
33. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet 2021; 53(10): 1415-24.
34. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet 2020; 52(7): 680-91.
35. Perez-Paramo YX, Watson CJW, Chen G, Thomas CE, Adams-Haduch J, Wang R, et al. Impact of genetic variants in the nicotine metabolism pathway on nicotine metabolite levels in smokers. Cancer Epidemiol Biomarkers Prev 2023; 32(1): 54-65.
36. Al-Khelaifi F, Diboun I, Donati F, Botrè F, Abraham D, Hingorani A, et al. Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance. Sci Rep 2019; 9(1): 19889.
37. Litvinov D, Selvarajan K, Garelnabi M, Brophy L, Parthasarathy S. Anti-atherosclerotic actions of azelaic acid, an end product of linoleic acid peroxidation, in mice. Atherosclerosis 2010; 209(2): 449-54.
38. Dennis JK, Sealock JM, Straub P, Lee YH, Hucks D, Actkins K, et al. Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. Genome Med 2021; 13(1): 6.
39. Manikpurage HD, Eslami A, Perrot N, Li Z, Couture C, Mathieu P, et al. Polygenic risk score for coronary artery disease improves the prediction of early-onset myocardial infarction and mortality in men. Circ Genomic Precis Med 2021; 14(6): e003452.