Journal of Diabetes and Its Complications 39 (2025) 109003

Contents lists available at ScienceDirect

Journal of Diabetes and Its Complications

journal homepage: www.elsevier.com/locate/jdiacomp

Type 2 diabetes and cardiovascular conditions prediction in individuals with metabolic syndrome-associated lipoprotein lipase gene (*LPL*) single nucleotide polymorphisms (SNPs)

Esphie Grace Fojas [a,*], Ahmad Haidery [a], Samina Naseeb [b], Roozbeh Naemi [c]

[a] University of Staffordshire, School of Health, Education, Policing and Sciences, Stoke-on-Trent, England, United Kingdom
[b] The University of Manchester, School of Biological Sciences, Faculty of Biology Medicine and Health, England, United Kingdom
[c] University of Salford, School of Health and Society, Manchester, England, United Kingdom

ARTICLE INFO

Keywords:
Metabolic syndrome
Cardiovascular disease
Type 2 diabetes
Obesity
BMI
Lipoprotein lipase

ABSTRACT

Objective: Metabolic syndrome (MetS) is predictive of increased risk of type 2 diabetes (T2D) and cardiovascular conditions (CVC). Lipoprotein lipase gene (*LPL*) single nucleotide polymorphisms (SNPs) may be of importance to the eventual diagnosis of T2D and CVC. This study aimed to predict the diagnosis of T2D and CVC amongst individuals with *LPL* SNPs rs268, rs11542065, rs116403115, rs118204057, rs118204061, rs144466625, and rs547644955.

Methods: This is a retrospective study using the UK Biobank data. Variables associated with MetS, T2D and CVC were selected from the data set. The total number of subjects in the cohort was 12,872 (mean age 56 years ± 8.1, 90.0 % were of British ethnicity, and 53.9 % were females). Logistic regression was used to assess whether the T2D and CVC can be predicted based on the presence of LPL SNPs and some of the clinical measures.

Results: Prediction models using clinical parameters showed good area under the curve (AUC) for prediction of T2D and CVC diagnosis (in receiver operating characteristic (ROC) analysis, area under the curve (AUC) = 0.959 for T2D, AUC = 0.772 for CVC). The addition of Polygenic Risk Scores (PRS/s) showed an improvement for diagnosis of both (AUC = 0.961 and 0.790 for TD and CVC, respectively). Further addition of SNPs showed more increase in AUC (AUC = 0.965 and 0.837 for T2D and CVC, respectively). The additive effect of the PRSs and *LPL* SNPs was more pronounced in the CVC than in the T2D model. The variant that had major significance for both T2D and CVC diagnoses was rs547644955 (AUC 1.0 and 0.910, respectively). The SNPs rs116403115 and rs118204057 both had an AUC of 1.0 for T2D diagnosis.

Conclusion: The prediction of T2D and CVC diagnoses with the use of clinically available factors may be enhanced with the addition of PRSs and SNPs, including *LPL* SNPs, which may have implications for stratified or personalised approaches for disease prevention or treatment.

1. Introduction

Metabolic syndrome (MetS) is a current global health concern and is predictive of increased risk of type 2 diabetes (T2D) and cardiovascular conditions (CVC).[1] Central to the cause of MetS and pathogenesis of T2D and CVC is obesity, particularly abdominal obesity, along with insulin resistance (IR), hyperglycemia, hypertension and dyslipidemia. There is a burgeoning rise of these inter-related diseases, and it has been estimated that about a quarter of the world's population, that is over a billion people, is affected by MetS.[2,3] The definition of MetS varies based on several criteria from various health authorities including World Health Organization (WHO; 1998), European Group of Insulin Resistance (EGIR; 1999), National Cholesterol Education Program Adult Treatment Panel III (NCEP:ATPIII; 2001), American Association of Clinical Endocrinologists (AACE; 2003), International Diabetes Federation (IDF; 2005), American Heart Association/National Heart, Lung, and Blood Institute (AHA/NHLBI; 2004) and a consensus definition incorporating AHA/NHLBI and IDF definitions (AHA/NHLBI/IDF; 2009).[4,5] Although the different criteria are all linked and in many aspects similar, evidently, there is a lack of universal MetS definition.

The individual components of MetS have been well-investigated with numerous reports on studies in IR, T2D, and CVC, for example. However,

* Corresponding author at: 7 Gateshead St, Athol Park, SA 5012, Australia.
E-mail address: esphiegrace.fojas@research.staffs.ac.uk (E.G. Fojas).

https://doi.org/10.1016/j.jdiacomp.2025.109003
Received 20 June 2024; Received in revised form 5 March 2025; Accepted 17 March 2025
Available online 24 March 2025

Table 1
Baseline information of the participants A. SNPs frequency distribution. B. Baseline characteristics. SNPs-single nucleotide polymorphisms, 2 SNPs-rs144466625; rs547644955 (n = 1), rs11542065; rs547644955 (n = 6), rs11542065; rs268 (n = 1), rs118204057; rs268 (n = 8), rs118204057; rs547644955 (n = 1), rs547644955; rs268 (n = 1). SD-standard deviation, BMI- body mass index, BP- blood pressure, HbA1c- glycated haemoglobin.

Variant ID	n
rs118204061	10
rs144466625	19
rs116403115	69
rs11542065	276
rs118204057	311
rs547644955	224
rs268	11,945
2 SNPs	18
N	12,872

A. SNPs frequency distribution

	N	Mean	SD
Weight (kg)	12,693	77.8	15.5
BMI (kg/m^2)	12,843	27.3	4.6
Waist circumference (cm)	12,856	89.9	13.2
Systolic BP (mmHg)	12,051	139.6	19.5
Diastolic BP (mmHg)	12,052	82.3	10.7
Glucose (mmol/L)	11,199	5.1	1.2
HbA1c (mmol/mol)	12,224	36	6.7

B. Baseline characteristics

publications in MetS as an entity are not at par in terms of quantity,[6] and more so on genetic studies although the heritability of MetS has been reported to be considerable at around 27 %.[7] Nevertheless, many studies are in agreement that IR is at the core in the pathogenesis of MetS, given the anabolic role of insulin to transport glucose in cells including adipocytes.[8,9] Increased free fatty acids (FFAs) levels arise when adipose tissues develop IR, promoting gluconeogenesis and very low-density lipoproteins (VLDLs) production.[10–13] The resulting impairment of lipoprotein homeostasis may be indicative of the role of lipoproteins in the development of obesity and MetS as well as the associated chronic diseases.[8]

The lipoprotein lipase (LPL), an enzyme which breaks down triglycerides (TG), is made by the instructions of the lipoprotein lipase gene (*LPL*).[14] CVC may arise from stored TG in fatty tissues which harden over time, and several reports also indicate that *LPL* variations may cause IR changes therefore, potentially resulting to obesity, MetS, and T2D.[4,14–16] When there is an irregularity with LPL function (i.e. an aberration of LPL gene in the form of SNP or mutation), this causes dysregulation of lipid metabolism and homeostasis which may lead to dyslipidemia. As an enzyme, when LPL hydrolyzes triglycerides, non-festered fatty acids (NEFA) and 2-monoacylglycerols are provided for many tissues- in the adipose tissue, NEFA is stored as triacylglycerol (TAG) via re-esterification; while in the muscles, NEFA is the major energy source, suggesting that LPL gene is the candidate gene for dyslipidemia.[17] Dyslipidemia may then lead to insulin resistance and/or pancreatic Beta cell apoptosis. Hence, *LPL* polymorphisms, including single nucleotide polymorphisms (SNPs) which are single nucleotide substitutions at a specific genomic locus location, may be of importance to the development of obesity and MetS. This may have repercussion in the eventual diagnosis of T2D and CVC, particularly amongst individuals with MetS.

Several LPL SNPs are of particular interest. In this study, those which were reported with conflicting findings (e.g. some reports indicate pathogenicity, others do not) for MetS or MetS-associated conditions in the National Institutes of Health (NIH) National Library Medicine, a web-based database, were used. These include rs268, rs11542065, rs116403115, rs118204057, rs118204061, rs144466625, and rs547644955 (seven SNPs). *LPL* SNPs have been evaluated to possibly aid in MetS diagnosis,[18] hence further studies on these SNPs in a large population may provide important additional information. These findings may help elucidate their role in the development of MetS as well as the diagnosis of T2D and CVC which can be of importance in the prevention and treatment of these diseases.

In this light, this study aims to predict the development of T2D or CVC as confirmed by definitive diagnosis amongst individuals with the afore-mentioned MetS-associated *LPL* SNPs.

2. Material and methods

The UK Biobank (UKB) data was solely used for this study (UKB reference for Research Ethics Committee (REC) approval 16/NW/0274). The UKB is a large-scale (involving approximately 500,000 individuals), long-term biobank in the United Kingdom that stores biospecimen and healthcare data, which are made available globally to advance scientific research. An online platform managed by DNANexus called the UKB Research Analysis Platform (RAP) was accessed from October–December 2023. The cohort of interest and variables associated with MetS, T2D, and CVC (sex, age, weight, BMI, waist circumference (WC), smoking and alcohol drinking status, physical activity, diet variation, blood pressure, cholesterol levels, HbA1c and glucose levels) as well as applicable parameters (diagnosis of T2D and CVC and relevant standard polygenic risk scores (PRS) – for T2D, cardiovascular disease (CVD), body mass index (BMI), glycated haemoglobin, coronary artery disease (CAD), atrial fibrillation (AF), high density lipoprotein (HDL), low density lipoprotein (LDL), and hypertension) were downloaded. Data were analysed using SPSS ver. 29.

CVC is defined for this study as any or combination of heart attack, angina, stroke, and high blood pressure; in the UKB data, this was collectively presented as presence of heart or vascular problems as diagnosed by a doctor. CVC is used as a surrogate marker of cardiovascular disease (CVD) in this study. T2D was identified in participants with ICD-10 diagnosis code E11. The predicted outcomes (dependent variables) are the diagnosis of T2D or CVC amongst individuals with the *LPL* SNPs of interest.

Average and standard deviation were calculated for each group for the selected parameters. No participant was excluded in the study due to participation withdrawal as per UKB's notice to UKB researchers. Participants with cancer (N = 1355) and those with other serious non-cancer medical condition or disability diagnosed by a doctor at baseline were excluded in the study. In the UKB, the latter was collectively presented in the UKB data based on a binary Yes/No information.

Direct logistic regression was performed to assess the impact of a set of predictor variables on the odds that the participants have been diagnosed with T2D or CVC at the time of recruitment. A total of four models have been assessed for both T2D and CVC diagnoses as outlined in the subheadings of the following sections.

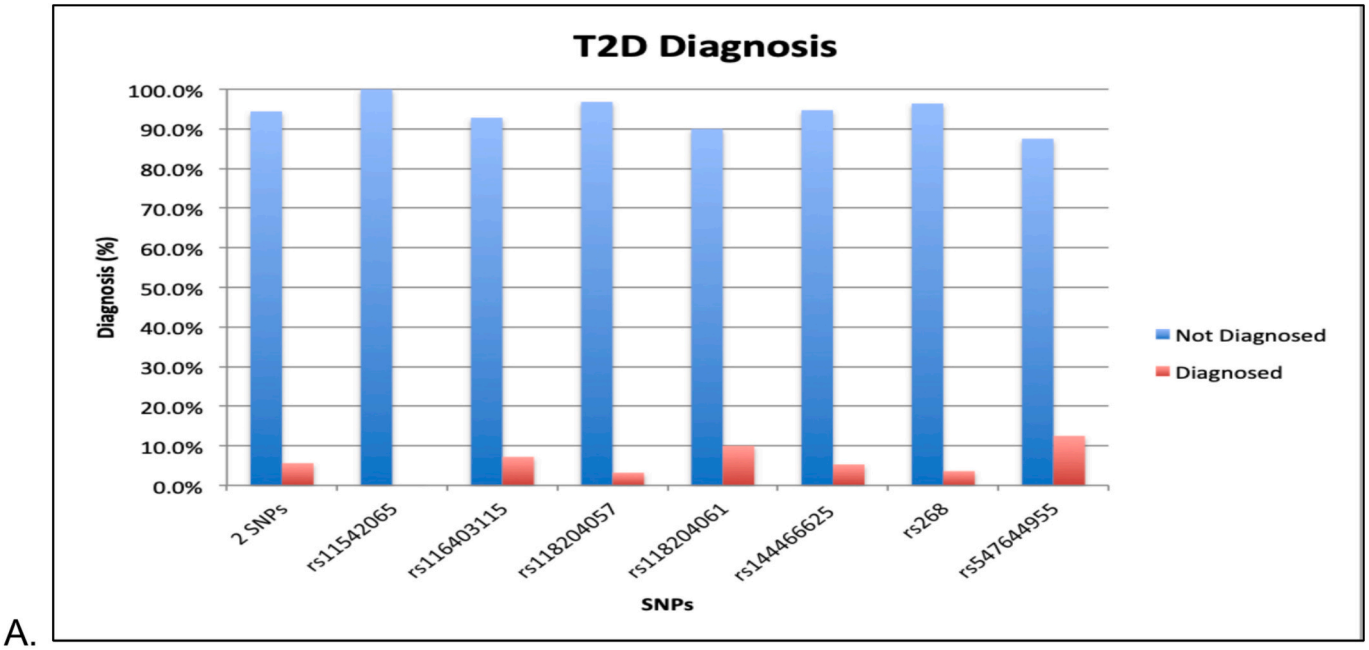
2.1. Prediction of T2D diagnosis

2.1.1. Prediction of T2D diagnosis by clinical parameters

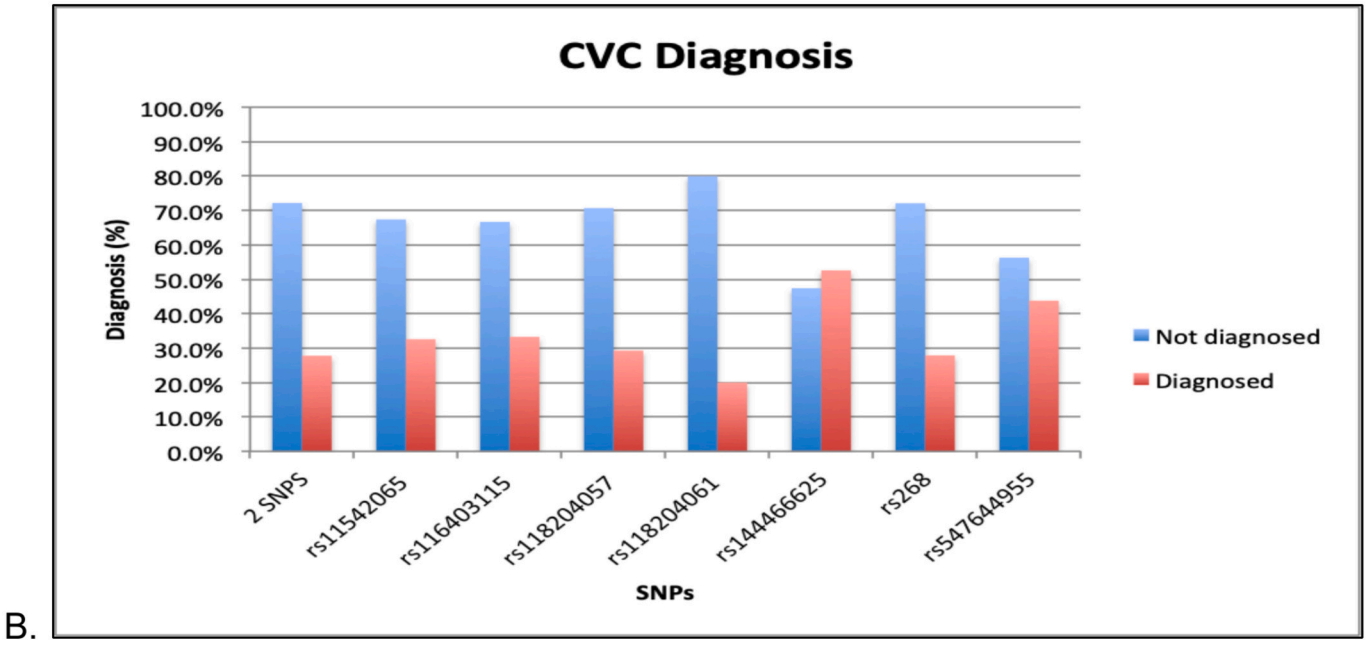
The model contained 15 independent variables normally accessible in clinical settings (age, sex, weight, height, BMI, waist circumference (WC), hip circumference HC), systolic BP, diastolic BP, number of days per week of moderate physical activity, diet variation, smoking status, alcohol drinking status, random blood glucose, HbA1c.

2.1.2. Prediction of T2D diagnosis by clinical parameters and T2D-associated Polygenic Risk Scores (PRS): Model A2

The model contained a total of 18 independent T2D-relevant variables including the above-stated clinical parameters (i.e. 15 variables) plus three T2D-relevant PRSs (PRS for T2D, BMI, and glycated



A.



B.

Fig. 1. Diagnosis of outcomes amongst individuals with the SNPs for: A. T2D, B. CVC showing percentage of diagnosed versus not diagnosed individuals. T2D- Type 2 diabetes, CVC- cardiovascular conditions, SNPs- single nucleotide polymorphisms, 2 SNPs- presence of 2 SNPs as defined in Table 1.

haemoglobin).

2.1.3. Prediction of T2D diagnosis by clinical parameters, T2D-associated PRS and SNPs: full model for T2D

Direct logistic regression was performed with the addition of the *LPL* SNP groups as a parameter to the above model (i.e. total 19 variables).

2.1.4. Prediction of T2D diagnosis by SNPs using Model A2

LPL SNP groups were filtered or selected as separate cases (i.e. Model A2 was used for individuals with rs268 only, and so on using other SNPs) and direct logistic regression was performed.

2.2. Prediction of CVC diagnosis

2.2.1. Prediction of CVC diagnosis by clinical parameters

The model contained 15 independent variables normally accessible in clinical settings, as with A.1.

2.2.2. Prediction of CVC diagnosis by clinical parameters and CVC-associated PRS: Model B2

The model contained a total of 22 independent CVC-relevant variables including the above-stated clinical parameters (i.e. 15 variables) plus seven CVC-relevant PRSs (PRS for BMI, CVD, atrial fibrillation, coronary artery disease (CAD), hypertension, LDL, HDL).

Table 2
Logistic regression outcome for T2D diagnosis using the full prediction model.

Variables	B	S.E.	Wald	df	Sig.	Exp(B)	95 % C.I. for EXP(B)	
							Lower	Upper
Age at recruitment	0.064	0.012	27.239	1	**<0.001**	1.066	1.041	1.092
Sex (1)	−0.055	0.265	0.044	1	0.834	0.946	0.563	1.591
Weight (kg)	−0.068	0.044	2.392	1	0.122	0.934	0.856	1.018
Height standing (cm)	0.088	0.048	3.352	1	0.067	1.092	0.994	1.2
BMI (kg/m^2)	0.355	0.13	7.491	1	**0.006**	1.426	1.106	1.84
Waist circumference (cm)	0.034	0.013	6.389	1	0.011	1.035	1.008	1.062
Hip circumference (cm)	−0.061	0.017	12.79	1	**<0.001**	0.941	0.91	0.973
Ever smoked			0.573	2	0.751			
Ever smoked (1)	1.055	1.494	0.498	1	0.48	2.871	0.154	53.695
Ever smoked (2)	1.004	1.492	0.453	1	0.501	2.73	0.147	50.798
Alcohol drinker status			1.787	3	0.618			
Alcohol drinker status (1)	−0.399	0.352	1.283	1	0.257	0.671	0.336	1.338
Alcohol drinker status (2)	−17.238	17,612.815	0	1	0.999	0	0	
Alcohol drinker status (3)	0.25	0.382	0.43	1	0.512	1.284	0.608	2.714
Systolic BP (mmHg)	0	0.005	0.007	1	0.932	1	0.989	1.01
Diastolic BP (mmHg)	−0.045	0.01	20.983	1	**<0.001**	0.956	0.938	0.975
Physical activity	−0.036	0.036	0.958	1	0.328	0.965	0.899	1.036
Variation in diet			2.461	3	0.482			
Variation in diet (1)	−0.832	0.88	0.894	1	0.344	0.435	0.078	2.441
Variation in diet (2)	−0.488	0.896	0.297	1	0.586	0.614	0.106	3.551
Variation in diet (3)	−0.737	0.874	0.711	1	0.399	0.479	0.086	2.655
Glucose (mmol/L)	−0.003	0.045	0.005	1	0.941	0.997	0.912	1.089
HbA1c (mmol/mol)	0.196	0.011	303.998	1	**<0.001**	1.216	1.19	1.243
Variant group			8.581	7	0.284			
Variant group (1)	1.438	3.692	0.152	1	0.697	4.211	0.003	5848.791
Variant group (2)	0.745	3.595	0.043	1	0.836	2.106	0.002	2416.766
Variant group (3)	−27.199	3198.5	0	1	0.993	0	0	.
Variant group (4)	−0.734	3.582	0.042	1	0.838	0.48	0	536.881
Variant group (5)	1.109	3.551	0.097	1	0.755	3.03	0.003	3194.619
Variant group (6)	0.353	3.527	0.01	1	0.92	1.424	0.001	1431.903
Variant group (7)	1.569	3.723	0.178	1	0.673	4.802	0.003	7087.11
Standard PRS for T2D	0.491	0.084	33.959	1	**<0.001**	1.634	1.385	1.927
Standard PRS for BMI	0.042	0.081	0.27	1	0.603	1.043	0.89	1.222
Standard PRS for HbA1c	−0.042	0.073	0.333	1	0.564	0.959	0.831	1.106
Constant	−28.068	8.919	9.902	1	0.002	0		

T2D- type 2 diabetes, BMI- body mass index, BP-blood pressure, HbA1c- glycated haemoglobin, PRS- polygenic risk score, df- degrees of freedom, Sig.- significance ($p < .05$ bolded to emphasize significance). Physical activity is defined as number of days/week of moderate physical activity 10+ minutes. SPSS coding: Sex- 0 for female, 1 for male; Ever smoked- 0 for No, 1 for Yes; Alcohol drinking status- 0 for never, 1 for previous, 2 for current (reference); Variation in diet- 0 for never/rarely, 1 for sometimes, 2 for often; Variant groups: 1-rs118204061 (reference), 2-rs1444466625, 3- rs116403115, 4- rs11542065, 5-rs118204057, 6- rs547644955, 7- rs268, 8- 2 SNPs.

2.2.3. Prediction of CVC diagnosis by clinical parameters, CVC-associated PRS and SNPs: full model for CVC

Direct logistic regression was performed with the addition of the *LPL* SNP groups as a parameter to the above model (i.e. total 23 variables).

2.2.4. Prediction of CVC diagnosis by SNPs using Model B2

LPL SNP groups were filtered or selected as separate cases and direct logistic was performed using Model B2.

3. Results

Normal distribution was found on test of data normality for all continuous variables. The total number of individuals in the UK Biobank for the seven specified SNPs was 17,386 when filtered individually, and 17,364 when filtered together, wherein the difference is attributed to participants having at least two SNPs. The total number of participants included in the study was 12,872; baseline information is presented in Table 1.

The highest number of frequency was for rs268 ($n = 11,945$). All 18 participants with multiple SNPs had 2 SNPs each, which were heterozygous for both SNPs. The majority in the cohort had heterozygous variations, and a small total of one hundred eleven (0.86 %) participants had homozygous variations involving rs116403115, rs11542065, and rs268 (1, 1, and 109 individuals, respectively). The percent of participants diagnosed with T2D or CVC per SNP groups are shown in Fig. 1.

Mean age of participants was 56 years ± 8.1; 53.9 % were females. 90.0 % were of British ethnicity. Mean weight and BMI at baseline were 77.8 ± 15.5 kg and 27.3 ± 4.6 kg/m^2, respectively.

Four hundred seventy-four (3.7 %) participants had T2D, and 3651 (28.4 %) had CVC. Most of the participants (93.3 %) were alcohol drinkers and majority (59.0 %) had history of smoking on study enrolment. Moderate exercise of at least 10 min for 5.4 ± 1.9 days per week was reported by most of the participants ($n = 12,681$). Variation in diet was recorded to be mostly sometimes (56.9 %) and never/rarely (34.8 %).

3.1. Prediction of T2D diagnosis

3.1.1. Prediction of T2D diagnosis by clinical parameters

The model containing all predictors (total 15 variables) was statistically significant, χ^2 (20, $N = 9668$) = 1399.3, $p < .001$. The accuracy, specificity, and sensitivity for the model were 97.1 %, 99.4 %, 38.5 %, respectively. The AUC in ROC analysis was 0.959 ($p < .001$).

3.1.2. Prediction of T2D diagnosis by clinical parameters and T2D-associated Polygenic Risk Scores (PRS): Model A2

The model containing all predictors (total 18 variables) was statistically significant, χ^2 (23, $N = 9623$) = 1427.2, $p < .001$. The model as a whole correctly classified 97.1 % of the cases, specificity was 99.4 %, sensitivity was 38.0 %. The AUC in ROC analysis was 0.961 ($p < .001$).

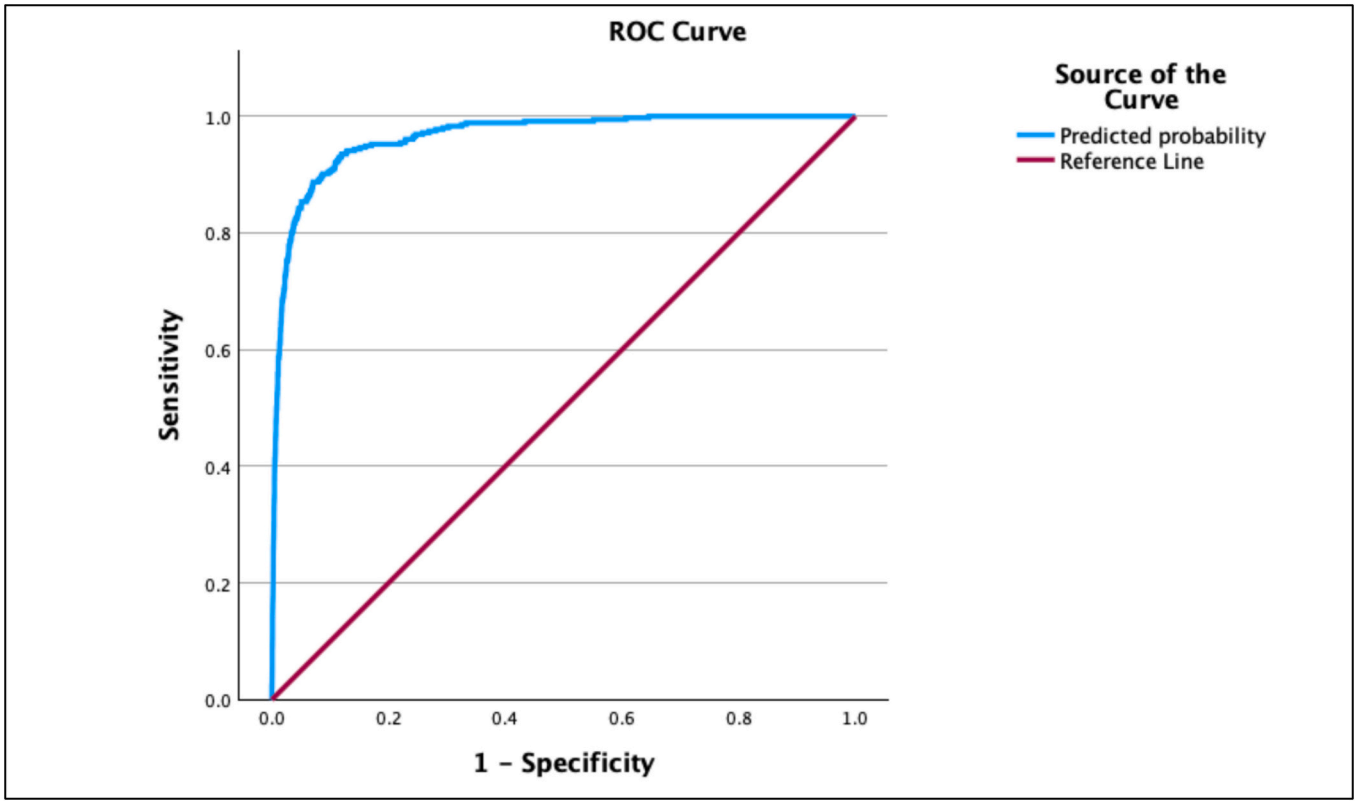


Fig. 2. ROC curve of the full model for prediction of T2D diagnosis (AUC = 0.965, $p < .05$).
ROC- receiver operating characteristic, T2D- type 2 diabetes, AUC- area under the curve.

3.1.3. Prediction of T2D diagnosis by clinical parameters, T2D-associated PRS and SNPs: full model for T2D

Addition of the SNPs as predictor (total 19 variables; χ^2 (30, N = 9623) = 1516.0, p < .001) correctly classified 97.3 % of the cases, specificity was the same at 99.4 %, and sensitivity increased to 42.5 %, a 4.5 % increase.

Six independent variables made a unique statistically significant contribution to the model, namely age, BMI, HP, diastolic BP, and standard PRS for T2D ($p < .05$; Table 2). The strongest predictor of being diagnosed with T2D in the participants with the *LPL* SNPs investigated in this study was the Standard PRS for T2D. This indicated that the odds were 1.6 times greater that the participants were diagnosed with T2D with per unit increase of PRS score, controlling for other factors in the model.

A ROC curve based on the full prediction model is presented in Fig. 2; the AUC was 0.965 ($p < .001$).

3.1.4. Prediction of T2D diagnosis by SNPs using Model A2

There were four SNPs (rs116403115, rs118204057, rs54764495, and rs268) with sufficient number of participants for analysis. On investigation of these SNPs, three SNPs had 100.0 % sensitivity and specificity – rs116403115, rs118204057, and rs54764495. The sensitivity for rs268, which had the highest number of N, was 42.5 %.

For the three SNPs with 100.0 % specificity, sensitivity, and accuracy (i.e. rs116403115, rs118204057, and rs54764495), the AUCs were 1.0 as expected ($p < .001$). The AUC for rs268 was 0.963 (p < .001).

3.2. Prediction of CVC diagnosis

3.2.1. Prediction of CVC diagnosis by clinical parameters

The model containing all predictors (total 15 variables) was statistically significant, χ^2 (20, N = 9668) = 1852.2, $p < .001$. The accuracy, specificity, and sensitivity for the model were 74.8 %, 91.4 %, 32.1 %, respectively. The AUC in ROC analysis was 0.772 ($p < .001$).

3.2.2. Prediction of CVC diagnosis by clinical parameters and CVC-associated PRS: Model B2

The model containing all predictors (total 22 variables) was statistically significant, χ^2 (27, N = 9623) = 2132.9, $p < .001$. The model as a whole correctly classified 75.8 % of the cases, specificity was 90.9 %, sensitivity was 37.1 %. The AUC in ROC analysis was 0.790 ($p < .001$).

3.2.3. Prediction of CVC diagnosis by clinical parameters, CVC-associated PRS and SNPs: full model for CVC

Addition of the SNPs as predictor (total 23 variables; χ^2 (34, N = 9623) = 2158.6, $p < .001$) correctly classified 75.9 % of the cases, specificity was the same at 90.9 %, and sensitivity increased to 37.5 %, a small 0.4 % increase.

Twelve independent variables made a unique statistically significant contribution to the model, namely age, WC, HC, alcohol drinker status, systolic BP, diastolic BP, variation in diet, HbA1c, variant group, and standard PRS for CVD, BMI, and hypertension ($p < .05$; Table 3); In reference to other SNPs, the odds of being diagnosed with CVC differs. Being a current alcohol drinker as well as the PRS for hypertension show high odds ration of 1.7 and 1.5, respectively, controlling for other factors in the model.

A ROC curve based on the full prediction model for CVC (AUC = 0.837, $p < .001$) is presented in Fig. 3. The AUC for the full model was 0.837 (p < .001).

3.2.4. Prediction of CVC diagnosis by SNPs using Model B2

On investigation of individuals per SNP groups with sufficient number of participants (four SNPs: rs11542065, rs118204057, rs54764495, and rs268), rs547644955 had the highest sensitivity at 75.9 %, specificity 83.1 %, and accuracy 80.9 % (ROC curve showing AUC = 0.910, $p < .001$, is shown in Fig. 3). The other 3 SNPs

Table 3
Logistic regression outcome for CVC diagnosis using the full prediction model.

Variables	B	S.E.	Wald	df	Sig.	Exp(B)	95 % C.I. for EXP(B)	
							Lower	Upper
Age at recruitment	0.07	0.004	311.141	1	**<0.001**	1.072	1.064	1.081
Sex (1)	−0.103	0.094	1.198	1	0.274	0.902	0.75	1.085
Weight (kg)	0.021	0.018	1.375	1	0.241	1.021	0.986	1.058
Height standing (cm)	−0.02	0.018	1.268	1	0.26	0.98	0.946	1.015
BMI(kg/m^2)	0.035	0.053	0.434	1	0.51	1.035	0.934	1.148
Waist circumference (cm)	0.03	0.005	36.574	1	**<0.001**	1.03	1.02	1.04
Hip circumference (cm)	−0.038	0.007	32.137	1	**<0.001**	0.962	0.95	0.975
Ever smoked			5.076	2	0.079			
Ever smoked (1)	0.247	0.468	0.278	1	0.598	1.28	0.512	3.201
Ever smoked (2)	0.363	0.467	0.604	1	0.437	1.438	0.576	3.59
Alcohol drinker status			12.934	3	0.005			
Alcohol drinker status (1)	0.064	0.135	0.225	1	0.635	1.066	0.818	1.39
Alcohol drinker status (2)	0.762	0.952	0.642	1	0.423	2.143	0.332	13.847
Alcohol drinker status (3)	0.518	0.148	12.226	1	**<0.001**	1.678	1.256	2.244
Systolic BP (mmHg)	0.016	0.002	69.259	1	**<0.001**	1.016	1.012	1.02
Diastolic BP (mmHg)	0.021	0.003	37.719	1	**<0.001**	1.021	1.014	1.028
Physical activity	−0.008	0.014	0.319	1	0.572	0.992	0.966	1.019
Variation in diet			22.046	3	**<0.001**			
Variation in diet (1)	−0.015	0.388	0.001	1	0.969	0.985	0.461	2.108
Variation in diet (2)	0.138	0.396	0.122	1	0.727	1.148	0.529	2.493
Variation in diet (3)	0.25	0.387	0.418	1	0.518	1.284	0.602	2.74
Glucose (mmol/L)	0.006	0.026	0.05	1	0.823	1.006	0.957	1.057
HbA1c (mmol/mol)	0.013	0.005	6.4	1	**0.011**	1.013	1.003	1.024
Variant group			25.564	7	**<0.001**			
Variant group (1)	2.924	1.268	5.314	1	0.021	18.617	1.549	223.696
Variant group (2)	2.238	1.169	3.663	1	0.056	9.377	0.948	92.786
Variant group (3)	1.966	1.134	3.005	1	0.083	7.143	0.773	65.966
Variant group (4)	1.518	1.131	1.803	1	0.179	4.565	0.498	41.865
Variant group (5)	2.276	1.138	3.998	1	0.046	9.734	1.046	90.597
Variant group (6)	1.586	1.12	2.006	1	0.157	4.884	0.544	43.853
Variant group (7)	1.353	1.331	1.034	1	0.309	3.871	0.285	52.593
Standard PRS for CVD	0.103	0.039	7.076	1	**0.008**	1.108	1.027	1.195
Standard PRS for BMI	−0.059	0.027	4.552	1	**0.033**	0.943	0.894	0.995
Standard PRS for AF	−0.007	0.029	0.066	1	0.797	0.993	0.938	1.05
Standard PRS for CAD	0.053	0.039	1.807	1	0.179	1.054	0.976	1.138
Standard PRS for hypertension	0.411	0.03	189.265	1	**<0.001**	1.508	1.422	1.599
Standard PRS for HDL cholesterol	−0.023	0.026	0.782	1	0.376	0.977	0.929	1.028
Standard PRS for LDL cholesterol	0.01	0.025	0.162	1	0.688	1.01	0.961	1.062
Constant	−9.443	3.238	8.507	1	0.004	0		

CVC- cardiovascular conditions, BMI- body mass index, BP-blood pressure, HbA1c- glycated haemoglobin, PRS- polygenic risk score, CVD-cardiovascular disease, AF-atrial fibrillation, CAD- coronary artery disease, HDL- high density lipoprotein, LDL- low density lipoprotein, df- degrees of freedom, Sig.- significance ($p < .05$ bolded to emphasize significance). Physical activity is defined as number of days/week of moderate physical activity 10+ minutes. SPSS coding: Sex- 0 for female, 1 for male; Ever smoked- 0 for No, 1 for Yes; Alcohol drinking status- 0 for never, 1 for previous, 2 for current (reference); Variation in diet- 0 for never/rarely, 1 for sometimes, 2 for often; Variant Groups: 1-rs118204061 (reference), 2-rs1444466625, 3- rs116403115, 4- rs11542065, 5-rs118204057, 6- rs547644955, 7- rs268, 8- 2 SNPs.

(rs11542065, rs118204057, and rs268) had low sensitivity (50.8, 33.8, and 36.9, respectively).

4. Discussion

Numerous research worldwide have now taken advantage of genome-wide association studies (GWAS) to discover genetic variants associated with disease traits, including obesity and related comorbidities. GWAS for MetS have been reported for various ethnic populations and from multiethnic backgrounds.[1,19,20] In 2011, a systematic review on the genetic variants associated with MetS has outlined the most studied SNPs linked with MetS; *LPL* was not included in this report.[21] A recent (2019) publication on MetS GWAS which has used the UK Biobank data has been published, which reported 80 novel independent loci; *LPL* SNP rs3844510 was included although not reported as novel.[22] The use of larger data sets such as in the latter as well as this study has been argued to be of significance particularly for linkage and candidate gene studies including MetS.[6] Nevertheless, current evidence suggests that the genetic risk factors for MetS are strongly connected with the components of MetS, including hyperglycemia and dyslipidemia.[23]

Prediction models have been trialled with the use of various parameters including risk factors to estimate the probability of T2D and/or CVD development in multiple studies using logistic regression and machine learning approaches.[24–26] These models may aid in formulating preventive measures for those who may be deemed at risk for developing the disease. In this study, the logistic regression model for T2D and CVC had high accuracy, specificity, and AUC in ROC analysis. Sensitivity was considerably low, except for rs547644955, rs116403115 and rs118204057. However, the ROC AUC has been established as the better assessment tool for medical diagnostic evaluation due to the arbitrary nature of specificity, sensitivity, and accuracy, which is deemed problematic.[27,28] The ROC AUC therefore better distinguishes between healthy versus diseased population,[29] and the models assessed in this study may be of value (Figs. 2 and 3), including the addition of SNPs particularly for individuals diagnosed with T2D (Results section B). Of note, the investigated SNPs may be further correlated with plasma cytokines, high-sensitivity C-reactive protein (hs-CRP), lipid profile, LDL particle size, and other relevant parameters for probability estimation of T2D and/or CVD diagnoses via prediction models in future studies, which will aid in better understanding of the multifactorial nature of these diseases.

The variant that had major significance for both T2D and CVC was rs547644955 (Results sections B and C). The other two variants with significance for T2D were rs116403115 and rs118204057 (Results section B). There appears to be no substantial publications for rs547644955 and rs116403115, therefore these findings may pave the way for greater

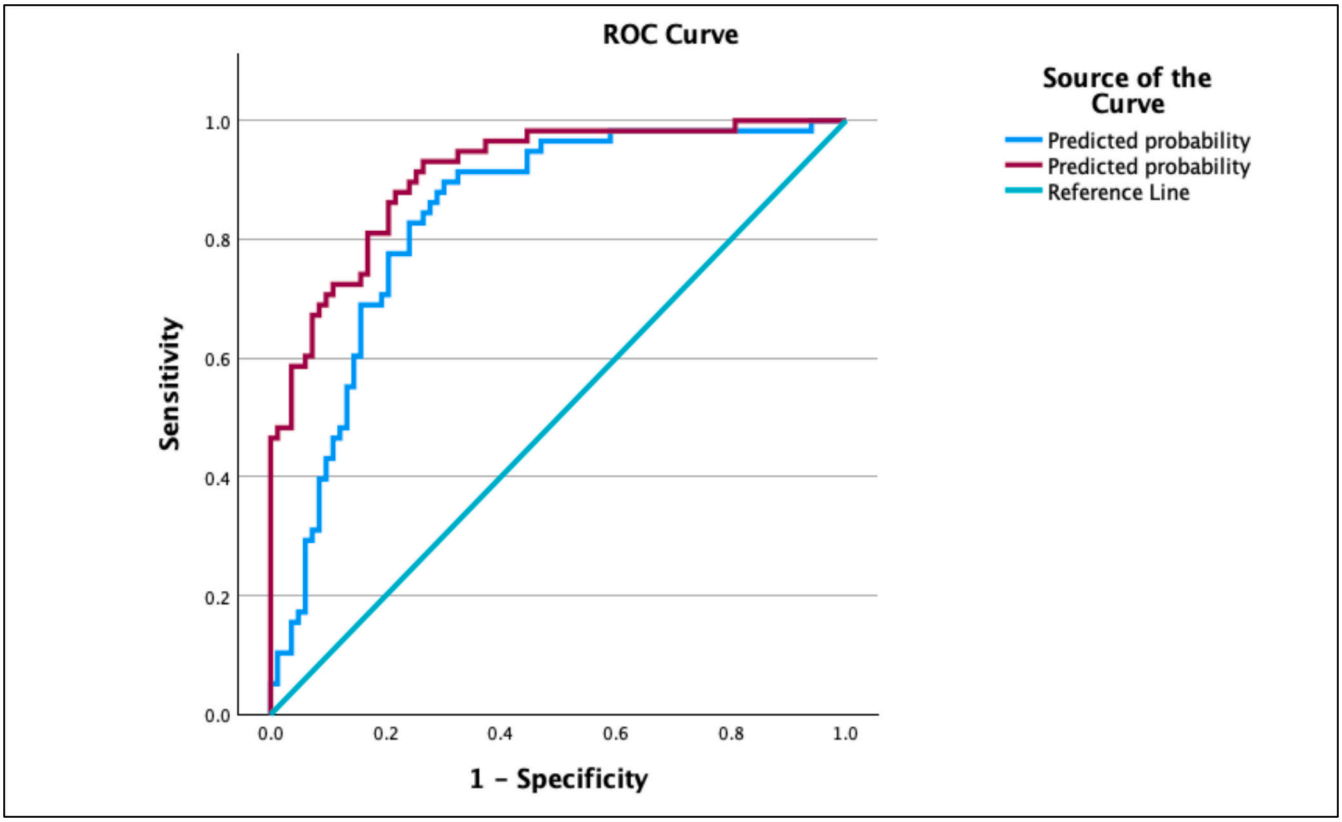


Fig. 3. ROC curve of the model for prediction of CVC diagnosis (full model, blue curve; AUC = 0.837, p < .001) and in individuals with rs547644955 (red curve; AUC = 0.910, $p < .001$).
ROC- receiver operating characteristic, CVC- cardiovascular conditions, AUC- area under the curve. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

understanding and clinical applications for individuals identified with these SNPs. There were a few reports directly associated with the variant rs118204057, including heritability in ethnic groups.[30–32] The variant rs268 was the most common in the cohort studied, and previous publications reported MetS-specific resuts.[33–35] Nonetheless, as the full model for both T2D and CVC diagnosis had good predictability based on the ROC analysis, these may be of considerable clinical and research use, particularly to obesity studies, overall.

Upon the exclusion of the genetic parameters on both full models, the decrease in AUC was marginal (0.959 versus 0.965 for the full model) for T2D (Results section B, D), while that for CVC may be considerable (0.772 versus 0.837 for the full model; Results section C, D). Although the clinical relevance may need further investigation, the availability of the fifteen variables may be readily accessible or easily obtainable in routine healthcare centres. In addition, it is imperative to note here that the parameters which showed statistically significant contribution to the full models, both for T2D (Table 2) and CVC (Table 3), would need to be examined more closely in succeeding studies for the investigated cohort or similar populations. On another note, T2D, although it relates to CVD with several similar risk factors, is known be a risk factor itself for the development of CVD, but does not seem to be true vice versa.[36–38] This was the primary consideration for the selection of PRSs for the *full* models in the T2D and CVC diagnosis prediction models. The *full* model for CVC diagnosis was trialled with the addition of PRSs for T2D and glycated haemoglobin, however the result of the AUC in ROC analysis did not differ (result not presented). The American Heart Association (AHA) has recently published (2022) a scientific statement regarding PRS for CVD as well as other related conditions such as T2D.[39] PRS is normally derived from single nucleotide variant effect sizes from GWAS then adjusted for linkage disequilibrium, and large biorepositories such as the UKB provides these data as what has been used in this study. As per summary of the AHA statement, the utility of PRS for CVD and associated disorders appears somehow different based on specific disease states as evidenced by various research. In CVD, CAD is the most studied form in terms of PRS research and its use is mainly geared towards pharmacological management.[40–42] In T2D, earlier studies point to similar utility of PRS with clinical factors, while more recent evidence suggests that PRS may be additive to the latter.[43–45] Yet other studies suggest unclear significance of T2D high-risk identification.[42,46] These findings may somehow be in accordance with the results of this research in terms of ambiguous usability of PRS addition to the prediction model. Although some research suggest that PRS for T2D may be of value for assessing response to sulfonylureas[47] as well as for glucose management,[48] therefore the clinical applications of PRS use may be worth pursuing in this age of advanced genomic evaluations.

A minor limitation of the study worth mentioning is the minimum age of participants that UKB includes (i.e. from 40 years old). In the recent age, more people get confirmed T2D or CVC diagnosis at a younger age (i.e. <40 years old). Although data for individual's age of diagnosis prior to UKB participation was specified in the database, age of diagnosis and modifiable risk factors in addition to genetics may have been further explored in longitudinal studies as possible, for instance. Nevertheless, research on younger population with the *LPL* SNPs evaluated in this study will certainly add value to the findings reported on this project.

5. Conclusions

In this study, models for prediction of T2D and CVC diagnosis have been explored using logistic regression from the UKB data. The addition

of genetic contribution enhanced the AUC values therefore the models better differentiate diagnosed versus non-diagnosed individuals. The additive effect of the *LPL* SNPs and relevant PRS was more pronounced in the CVC than in the T2D model. The variant that had major significance for both T2D and CVC diagnoses was rs547644955, with an AUC of 1.0 and 0.910, respectively. The SNPs rs116403115 and rs118204057 both had an AUC of 1.0 for T2D diagnosis. Additional research is required to further investigate the effects of these *LPL* SNPs in the development of MetS and other obesity-related diseases, which may have impact for stratified or personalised approaches for disease prevention or treatment.

CRediT authorship contribution statement

Esphie Grace Fojas: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Ahmad Haidery:** Writing – review & editing. **Samina Naseeb:** Writing – review & editing. **Roozbeh Naemi:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Acknowledgment

This research has been conducted using the UK Biobank Resource under Application Number 77577.

References

1. Oh SW, Lee JE, Shin E, Kwon H, Choe EK, Choi SY, et al. Genome-wide association study of metabolic syndrome in Korean populations. *PloS One*. 2020;15, e0227357.
2. Grundy SM. Metabolic syndrome pandemic. *Arterioscler Thromb Vasc Biol*. 2008;28:629–636.
3. Saklayen MG. The global epidemic of the metabolic syndrome. *Curr Hypertens Rep*. 2018;20:12.
4. Huang PL. A comprehensive definition for metabolic syndrome. *Dis Model Mech*. 2009;2:231–237.
5. Kassi E, Pervanidou P, Kaltsas G, Chrousos G. Metabolic syndrome: definitions and controversies. *BMC Med*. 2011;9:48.
6. Monda KL, North KE, Hunt SC, Rao DC, Province MA, Kraja AT. The genetics of obesity and the metabolic syndrome. *Endocr Metab Immune Disord Drug Targets*. 2010;10:86–108.
7. O'Neill S, O'Driscoll L. Metabolic syndrome: a closer look at the growing epidemic and its associated pathologies. *Obes Rev*. 2015;16:1–12.
8. Fahed G, Aoun L, Bou Zerdan M, Allam S, Bou Zerdan M, Bouferraa Y, et al. Metabolic syndrome: updates on pathophysiology and management in 2021. *Int J Mol Sci*. 2022;23.
9. Saltiel AR. Insulin signaling in health and disease. *J Clin Invest*. 2021;131.
10. Boden G, Shulman GI. Free fatty acids in obesity and type 2 diabetes: defining their role in the development of insulin resistance and beta-cell dysfunction. *Eur J Clin Invest*. 2002;32:14–23.
11. Griffin ME, Marcucci MJ, Cline GW, Bell K, Barucci N, Lee D, et al. Free fatty acid-induced insulin resistance is associated with activation of protein kinase C theta and alterations in the insulin signaling cascade. *Diabetes*. 1999;48:1270–1274.
12. Murakami T, Michelagnoli S, Longhi R, Gianfranceschi G, Pazzucconi F, Calabresi L, et al. Triglycerides are major determinants of cholesterol esterification/transfer and HDL remodeling in human plasma. *Arterioscler Thromb Vasc Biol*. 1995;15:1819–1828.
13. Unger RH, Zhou YT. Lipotoxicity of beta-cells in obesity and in other causes of fatty acid spillover. *Diabetes*. 2001;50:S118–S121.
14. Pirahanchi Y, Anoruo M, Sharma S. *Biochemistry, Lipoprotein Lipase*. StatPearls Publishing LLC; 2023 [StatPearls. Treasure Island (FL) ineligible companies. Disclosure: McDamian Anoruo declares no relevant financial relationships with ineligible companies. Disclosure: Sandeep Sharma declares no relevant financial relationships with ineligible companies.: StatPearls Publishing. Copyright © 2023].
15. Goodarzi MO, Guo X, Taylor KD, Quiñones MJ, Saad MF, Yang H, et al. Lipoprotein lipase is a gene for insulin resistance in Mexican Americans. *Diabetes*. 2004;53:214–220.
16. Kumari A, Kristensen KK, Ploug M, Winther AL. The importance of lipoprotein lipase regulation in atherosclerosis. *Biomedicines*. 2021;9.
17. Xie C, Wang ZC, Liu XF, Yang MS. The common biological basis for common complex diseases: evidence from lipoprotein lipase gene. *Eur J Hum Genet*. 2010;18:3–7.
18. Kang SW, Kim SK, Kim YS, Park MS. Risk prediction of the metabolic syndrome using TyG Index and SNPs: a 10-year longitudinal prospective cohort study. *Mol Cell Biochem*. 2023;478:39–45.
19. Moon S, Lee Y, Won S, Lee J. Multiple genotype-phenotype association study reveals intronic variant pair on SIDT2 associated with metabolic syndrome in a Korean population. *Hum Genomics*. 2018;12:48.
20. Tekola-Ayele F, Doumatey AP, Shriner D, Bentley AR, Chen G, Zhou J, et al. Genome-wide association study identifies African-ancestry specific variants for metabolic syndrome. *Mol Genet Metab*. 2015;116:305–313.
21. Povel CM, Boer JM, Reiling E, Feskens EJ. Genetic variants and the metabolic syndrome: a systematic review. *Obes Rev*. 2011;12:952–967.
22. Lind L. Genome-wide association study of the metabolic syndrome in UK Biobank. *Metab Syndr Relat Disord*. 2019;17:505–511.
23. Taylor JY, Kraja AT, de Las FL, Stanfill AG, Clark A, Cashion A. An overview of the genomics of metabolic syndrome. *J Nurs Scholarsh*. 2013;45:52–59.
24. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Med Inform Decis Mak*. 2019;19:211.
25. Edlitz Y, Segal E. Prediction of type 2 diabetes mellitus onset using logistic regression-based scorecards. *Elife*. 2022:11.
26. Joshi RD, Dhakal CK. Predicting type 2 diabetes using logistic regression and machine learning approaches. *Int J Environ Res Public Health*. 2021;18.
27. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013;4:627–635.
28. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988;240:1285–1293.
29. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8:283–298.
30. Gagné C, Brun LD, Julien P, Moorjani S, Lupien PJ. Primary lipoprotein-lipase-activity deficiency: clinical investigation of a French Canadian population. *Cmaj*. 1989;140:405–411.
31. Henderson HE, Hassan F, Berger GM, Hayden MR. The lipoprotein lipase Gly188——Glu mutation in South Africans of Indian descent: evidence suggesting common origins and an increased frequency. *J Med Genet*. 1992;29:119–122.
32. Paulweber B, Wiebusch H, Miesenboeck G, Funke H, Assmann G, Hoelzl B, et al. Molecular basis of lipoprotein lipase deficiency in two Austrian families with type I hyperlipoproteinemia. *Atherosclerosis*. 1991;86:239–250.
33. Franceschini N, Carty C, Bůzková P, Reiner AP, Garrett T, Lin Y, et al. Association of genetic variants and incident coronary heart disease in multiethnic cohorts: the PAGE study. *Circ Cardiovasc Genet*. 2011;4:661–672.
34. Nejati M, Atlasi MA, Karimian M, Nikzad H, Tameh AA. Lipoprotein lipase gene polymorphisms as risk factors for stroke: a computational and meta-analysis. *Iran J Basic Med Sci*. 2018;21:701–708.
35. Sagoo GS, Tatt I, Salanti G, Butterworth AS, Sarwar N, van Maarle M, et al. Seven lipoprotein lipase gene polymorphisms, lipid fractions, and coronary disease: a HuGE association review and meta-analysis. *Am J Epidemiol*. 2008;168:1233–1246.
36. Dal Canto E, Ceriello A, Rydén L, Ferrini M, Hansen TB, Schnell O, et al. Diabetes as a cardiovascular risk factor: an overview of global trends of macro and micro vascular complications. *Eur J Prev Cardiol*. 2019;26:25–32.
37. Einarson TR, Acs A, Ludwig C, Panton UH. Economic burden of cardiovascular disease in type 2 diabetes: a systematic review. *Value Health*. 2018;21:881–890.
38. Kelsey MD, Nelson AJ, Green JB, Granger CB, Peterson ED, McGuire DK, et al. Guidelines for cardiovascular risk reduction in patients with type 2 diabetes: JACC guideline comparison. *J Am Coll Cardiol*. 2022;79:1849–1857.
39. O'Sullivan JW, Raghavan S, Marquez-Luna C, Luzum JA, Damrauer SM, Ashley EA, et al. Polygenic risk scores for cardiovascular disease: a scientific statement from the American Heart Association. *Circulation*. 2022;146:e93–e118.
40. Damask A, Steg PG, Schwartz GG, Szarek M, Hagström E, Badimon L, et al. Patients with high genome-wide polygenic risk scores for coronary artery disease may receive greater clinical benefit from alirocumab treatment in the ODYSSEY OUTCOMES trial. *Circulation*. 2020;141:624–636.
41. Khera AV, Emdin CA, Drake I, Natarajan P, Bick AG, Cook NR, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375:2349–2358.
42. Said MA, Verweij N, van der Harst P. Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK Biobank study. *JAMA Cardiol*. 2018;3:693–702.
43. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*. 2020;26:549–557.
44. 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;359:2208–2219.
45. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *Bmj*. 2010;340, b4838.
46. Hivert MF, Jablonski KA, Perreault L, Saxena R, McAteer JB, Franks PW, et al. Updated genetic score based on 34 confirmed type 2 diabetes Loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. *Diabetes*. 2011;60:1340–1348.
47. Li JH, Szczerbinski L, Dawed AY, Kaur V, Todd JN, Pearson ER, et al. A polygenic score for type 2 diabetes risk is associated with both the acute and sustained response to sulfonylureas. *Diabetes*. 2021;70:293–300.
48. Shah HS, Gao H, Morieri ML, Skupien J, Marvel S, Paré G, et al. Genetic predictors of cardiovascular mortality during intensive glycemic control in type 2 diabetes: findings from the ACCORD clinical trial. *Diabetes Care*. 2016;39:1915–1924.