



## Research article

## Decreased plasma miR-140-3p is associated with coronary artery disease

Pei Mo<sup>a,1</sup>, Chao-Wei Tian<sup>b,1</sup>, Qiqi Li<sup>c</sup>, Mo Teng<sup>d</sup>, Lei Fang<sup>a</sup>, Yujuan Xiong<sup>e,\*\*</sup>, Benrong Liu<sup>a,\*</sup>

<sup>a</sup> Department of Cardiology, Guangzhou Institute of Cardiovascular Disease, Guangdong Key Laboratory of Vascular Diseases, State Key Laboratory of Respiratory Disease, The Second Affiliated Hospital, Guangzhou Medical University, Guangzhou, 510260, China

<sup>b</sup> Department of General Practice, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, 510260, China

<sup>c</sup> Department of Medical Imaging, Second Clinical College, Guangzhou Medical University, Guangzhou, 510260, China

<sup>d</sup> Department of Obstetrics, The Second Affiliated Hospital, Guangzhou Medical University, Guangzhou, 510260, China

<sup>e</sup> Department of Laboratory Medicine, Panyu Hospital of Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou, 511400, China

## ARTICLE INFO

## Keywords:

miRNA  
miR-140-3p  
Machine learning  
Coronary artery disease  
Biomarker

## ABSTRACT

**Background:** Although many circulating miRNAs (c-miRNAs) are associated with coronary artery disease (CAD), they are far from being the biomarker for CAD diagnosis or risk prediction. Therefore, novel c-miRNAs discovery and validation are still required, especially evaluating their prediction capacity.

**Objectives:** Identify novel CAD-related c-miRNAs and evaluate its risk prediction capacity for CAD. **Methods:** miRNAs associated with CAD were preliminarily investigated in three paired samples representing pre-CAD stage and CAD stage of three female individuals using the Applied Biosystems miRNA TaqMan® Low-Density Array (TLDA). Then, the candidate miRNAs were further verified in an independent case-control study including 129 CAD patients and 76 controls, and their potential practical value in prediction for CAD was evaluated using a machine learning (ML) algorithm. The accuracy of classification and prediction was assessed with the area under the receiver operating characteristic curve (AUC).

**Results:** TLDA analysis shows that miR-140-3p decreased significantly in CAD-stage (FC = -3.01,  $P = 0.007$ ). Further study shows that miR-140-3p was significantly lower in CAD group [1.26 (0.68, 2.01)] than in control group [2.07 (1.19, 3.21)] ( $P < 0.001$ ) and independently associated with CAD ( $P < 0.001$ ). The addition of miR-140-3p to the variables including smoking history, HDL-c, and APOA1 improved the accuracy of classification by logistic regression and of prediction for CAD by ML models. The ML models built with miR-140-3p and HDL-c, respectively, had a similar prediction accuracy. The feature importance of miR-140-3p and HDL-c in the ML models was also similar. Decision curve analysis showed that miR-140-3p and HDL-c had almost identical net benefits.

**Conclusion:** Reduced levels of miR-140-3p is linked to CAD, and it is possible to use the plasma level of miR-140-3p as a means of evaluating the risk of CAD.

\* Corresponding author. Guangzhou Institute of Cardiovascular Disease, Guangdong Key Laboratory of Vascular Diseases, the Second Affiliated Hospital, Guangzhou Medical University, 250 Changgang Dong Road, Guangzhou, 510260, China.

\*\* Corresponding author. Department of Laboratory Medicine, Panyu Hospital of Chinese Medicine, Guangzhou University of Chinese Medicine, 93 Qiaodong Road, Guangzhou, 511400, China.

E-mail addresses: [yujuanxiong@gzucm.edu.cn](mailto:yujuanxiong@gzucm.edu.cn) (Y. Xiong), [liubenrong@gzhmu.edu.cn](mailto:liubenrong@gzhmu.edu.cn) (B. Liu).

<sup>1</sup> These authors contributed equally.

<https://doi.org/10.1016/j.heliyon.2024.e26960>

Received 7 December 2022; Received in revised form 19 February 2024; Accepted 22 February 2024

Available online 25 February 2024

2405-8440/© 2024 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Abbreviations

CAD	coronary artery disease
CVD	cardiovascular disease
ML:	machine learning
HDL-c:	high-density lipoprotein cholesterol
LDL-c:	low-density lipoprotein cholesterol
ApoA1	apolipoprotein A
ApoB	apolipoprotein B
CHOL:	total Cholesterol
TG	triglycerides
GLU	glucose
LR	Logistic Regression
RFC	Random Forest Classifier
GBC	Gradient Boosting Classifier
ROC:	receiver operating characteristic curve
AUC	area under the receiver operating characteristic curve
LASSO	least absolute shrinkage and selection operator
miRNA-TLDA	Applied Biosystems miRNA TaqMan® low density array

## 1. Introduction

Coronary artery disease (CAD) has been proven to be the major cause of death in both the developed and developing countries [1]. Early warning and intervention for CAD high-risk groups are of great significance in delaying the onset of CAD and reducing mortality. Identification of novel biomarkers is required for advancing the early prevention and treatment of CAD. MicroRNAs (miRNAs) are proven to be linked with many disease-status and are considered to be the potential biomarker in wide use for diagnosis, prognosis, therapy targets, and risk prediction [2–4]. Circulating miRNAs are stable in blood, urine, and other body fluids and are always concerned in the field of biomarker research due to the easy acquisition of samples, and have been widely studied in elderly-related diseases, especially in cancer, nervous disease, and cardiovascular diseases [5,6]. An altered miRNA expression profile is linked to cardiovascular pathogenesis including CAD, and may have utility as diagnostic and prognostic biomarkers [7,8]. Systematic analysis of several studies shows that miR-21, miR-133, and miR-499 have potential diagnostic value for acute coronary syndrome or stable CAD [7,9]. Myocardial injury after acute myocardial infarction increases the circulating levels of miRNAs (eg, miR-1, miR-133, miR-499, miR-208) in cardiac muscle, but in CAD or diabetes patients, the peripheral circulating level of miRNAs (miR-26) enriched in endothelial cells is reduced [10]. Combining traditional cardiovascular risk factors (TCRFs) and circulatory miRNAs can improve the prediction performance of the cardiovascular risk model. For example, when BNP (brain natriuretic peptide) is used alone or in combination with circulating miR-30c, miR-221, miR-328, miR-146a, or miR-375 as part of the diagnostic panel, the diagnostic performance of heart failure is significantly improved compared with BNP alone [11]. Although miRNAs have great potential as biomarkers for clinical diagnosis or prognostic evaluation of cardiovascular diseases (CVD), the results reported in relevant studies have rarely been further verified, thus hindering their clinical application [12,13]. Among dozens of circulating miRNAs associated with CVD, the number of miRNAs that can be used as diagnostic markers of CVD is limited [14]. Plasma levels of many miRNAs (miR-1, miR-133a, miR-208a/b, miR-499, and so on) involved in heart development, function, and damage repair have been reported to be associated with heart failure and acute myocardial infarction, these miRNAs may have potential value for the diagnosis of CVD [14]. MiRNAs related to vascular endothelial function or inflammation play key roles in the pathological process of atherosclerotic CAD [15, 16]. Changes in circulating miRNAs are reported to be associated with CAD, but it is seldom reported that they are used for early diagnosis or risk prediction. Therefore, it is necessary to find and validate miRNAs that can be used as circulating biomarkers to predict CAD risk.

Although numerous risk factors for CAD have been identified and studied, including circulating miRNAs, blood lipids, cholesterol, glucose, and inflammatory factors [17–20], utilizing these factors to predict CAD risk is not yet common. This is due to the large variability of some of these factors in different individuals, particularly inflammatory factors, which have been shown to be associated with CAD risk but are also sensitive to environmental factors such as diet and infection. Additionally, the limitations of the methods used in previous studies have led to low reliability in evaluating these factors for predicting CAD risks. Therefore, adopting new prediction or evaluation methods may yield the discovery of new plasma biomarkers related to CAD. The application of Machine learning (ML) algorithms in medicine improves the accuracy and applicability of biomarkers for early diagnosis, risk prediction, and prognosis evaluation of diseases [21,22]. New disease-related molecular markers can also be found by the ML algorithm, which has higher reliability than traditional statistical methods [23–25]. For example, Saberi-Karimian et al. [26] using decision tree analysis revealed that serum HDL functionality was a more important variable than HDL-C level in predicting patients with hypertension at baseline. Therefore, the application of new analytical methods including ML may be powerful in the discovery of circulating biomarkers.

In the present study, we aimed to discover novel circulating miRNAs associated with CAD risk, compare their ability with that of

TCRFs in prediction for CAD risk, and evaluate whether it improves the prediction accuracy for CAD risk when it was used in combination with TCRFs to construct ML models. We used the AB-miRNA-TLDA (Applied Biosystems miRNA TaqMan® Low-Density Array) analysis to detect the circulating miRNA expression profiles of the patients in pre-CAD-stage (stage of 3–5 years before being diagnosed with CAD) and CAD-stage (stage of diagnosed with CAD). The plasma level of miR-140-3p was found to be negatively associated with the progress of CAD and was further validated in a case-control study. Decreased plasma level of miR-140-3p was found to be associated with CAD risk. The role of miR-140-3p in the prediction for CAD risk was proven as important as that of HDL-c by different analyses such as the ML model, Least Absolute Shrinkage and Selection Operator (LASSO) regression, decision curve, and binary logistic regression.

## 2. Materials and methods

### 2.1. Samples for AB-miRNA-TLDA

Three paired plasma samples from three female individuals in pre-CAD-stage and CAD-stage were extracted from the sample bank. Total RNA was extracted from the plasma samples and used to detect the expression profile of circulating miRNAs using AB-miRNA-TLDA. The expression level of miRNAs may be affected by age which is a clear risk factor for CAD. Therefore, we specially selected three patients at different age range for miRNA-TLDA analysis (Table S1), to a certain extent, this reduced the age difference effect.

### 2.2. AB-miRNA-TLDA

TLDA is an approach to quantifying miRNA expression based on the Megaplex reverse transcription format of the stem-loop primer-based real-time quantitative polymerase chain reaction (RT-qPCR), which is more accurate and sensitive than traditional microarray based on hybridization with oligonucleotide probe hybridization [27]. The TLDA analysis was performed by CapitalBio Corporation (Beijing, China) with Megaplex™ Pools which are designed to detect and quantitate up to 380 miRNAs per pool using Applied Biosystems real-time instruments. Briefly, The TaqMan microRNA reverse transcription reaction mixture in which the Megaplex™ RT Primers included was applied to synthesize single-stranded cDNA from total RNA samples. Then, the Pre-amplification Reactions and the Real-time PCR reactions were performed stringently according to the manufacturer's protocol for Megaplex™ Pools.

### 2.3. Identification of differential miRNAs

NormFinder is an algorithm for identifying the optimal normalization gene among a set of candidates [28]. The NormFinder analysis showed that miR-151a-3p and miR-27b had a higher expression level in the plasma and were more stable than others. Thus, miR-151a-3p and miR-27b were selected as the reference gene to normalize the expression of all miRNAs tested with miRNA-TLDA analysis, as follows: the delta Ct1 was computed by the Ct of the target miRNAs minus the Ct of miR-151a-3p, and the delta Ct2 was computed by the Ct of the target miRNAs minus the Ct of miR-27b. Then the mean delta Ct was calculated by averaging the delta Ct1 and delta Ct2. The fold change relative to miR-151a-3p and miR-27b was calculated according to  $2^{-(\text{mean delta Ct})}$ , which was used to indicate the relative plasma level of the miRNAs. Paired *t*-test was used to detect the significant difference between the two groups for each miRNA,  $P < 0.05$  was considered to be statistically significant.

### 2.4. Participants for subsequent case-control study

The inpatients who were hospitalized from 2010 to 2017 in the department of cardiovascular disease, the Second Affiliated Hospital of Guangzhou Medical University were included in CAD and control groups, respectively, according to the criteria: 1) CAD group: medical records show that patients diagnosed as CAD by coronary angiography, that was, patients whose coronary angiography showed a greater than 50% stenosis of at least one vessel in the left main coronary artery, left anterior descending artery (including the main diagonal branch), circumflex artery (including the main marginal branch) and right coronary artery (including the posterior descending artery or left ventricular posterior branch); 2) control group: men aged  $\geq 50$  and women aged  $\geq 55$ , who were not diagnosed with CAD by coronary angiography, or had no symptoms such as chest tightness and chest pain, and no myocardial ischemia suggested by cardiac ultrasound diagnosis. Both groups excluded the patients with multiple organ failure, malignant tumors, and autoimmune deficiency. Finally, 129 CAD patients and 76 controls were included and their plasma samples preserved in the sample bank were used for validating the CAD-related miRNAs. Information of the laboratory indicators such as HDL-c, low-density lipoprotein cholesterol (LDL-c), triglycerides (TG), total cholesterol (CHOL), glucose (GLU), apolipoprotein A1 (ApoA1), and apolipoprotein B (ApoB), and the clinical information such as gender, age, smoking history (IsSmoking), hypertension, hyperlipidemia, and diabetes were acquired from the patient's electronic medical record.

### 2.5. Quantitative detection of the plasma miRNAs

Total miRNAs were extracted from 200  $\mu$ l of plasma by using miRcute serum/plasma miRNA extraction and separation Kit (DP503). The first strand cDNA of miRNA was synthesized by using miRcute Plus miRNA First-Strand cDNA Kit (KR211) which applies the method of adding adenine nucleotide to the 3' of miRNAs before reverse transcribing. The plasma level of the miR-140-3p and miR-151a-3p was detected by real-time fluorescent quantitative PCR (qPCR) using miRcute Plus miRNA qPCR Kit (FP411). miR-151a-3p

was used as a reference to normalize the expression level of miR-140-3p. The plasma level of miR-140-3p was represented with the multiples relative to miR-151a-3p, that was calculated according to the formula:  $2^{-(\text{Ct}[\text{miR}-140-3\text{p}]-\text{Ct}[\text{miR}-151\text{a}-3\text{p}-1])}$ .

## 2.6. Evaluation of the ability of factors in prediction for CAD risk

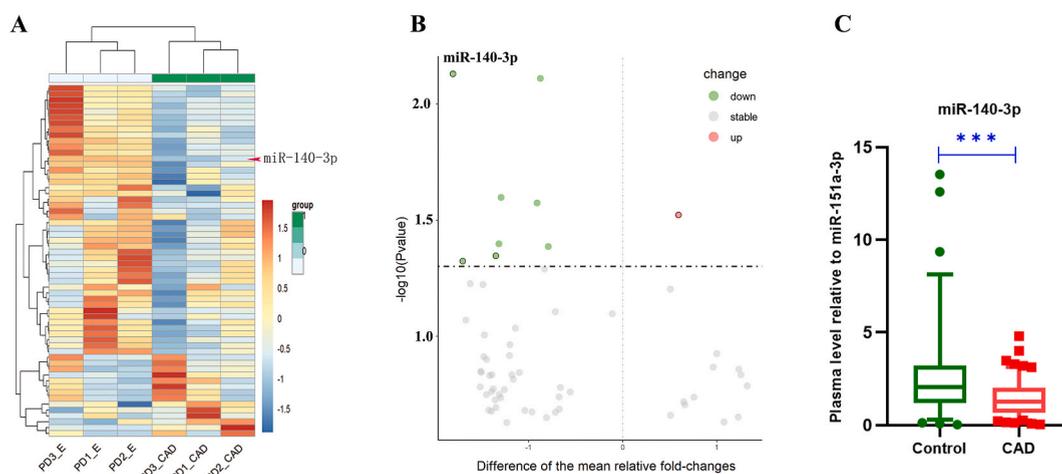
The ability of the factors statistically related to CAD in prediction for CAD risk was evaluated by constructing machine learning (ML) models with these factors in combination or separately. The dataset including the response variable “CAD” and the independent variables such as ‘Gender’, ‘Smoking’, ‘Drinking’, ‘Hypertension’, ‘Diabetes’, ‘Hyperlipidemia’, ‘Age’, ‘GLU’, ‘TG’, ‘CHOL’, ‘HDL-c’, ‘LDL-c’, ‘APOA1’, ‘APOB’, ‘miR-140-3p’ was analyzed by binary logistic regression (SPSS v.19.0.0). The variables remained in the final logistic regression model were used in combination or separately to build ML model. Dataset splitting was performed with the ‘train\_test\_split’ module in the python package of ‘sklearn.model\_selection’, and the training dataset was used to construct ML models using the ‘LogisticRegression classifier (LRC)’ in ‘sklearn.linear\_model’, ‘RandomForestClassifier (RFC)’ and ‘GradientBoostingClassifier (GBC)’ in ‘sklearn.ensemble’, respectively. Five-fold cross-validation was performed for the model construction. The test dataset was used to evaluate the performance of the ML models by calculating the area under the receiver operating characteristic curve (AUC). Average feature importance was computed for 20 RFCs and for 20 GBCs that were built with the training dataset split by tuning the values of random\_state.

## 2.7. Evaluation of the importance of the variables associated with CAD

LASSO regression is characterized by variable selection and complexity adjustment while fitting the generalized linear model. ‘Glmnet’, an R package, was used for LASSO regression analysis. The hyperparameters were set as family = ‘binomial’, type.measure = “AUC”, keep = true, nfolds = 5. Decision curve analysis (DCA) can be used to evaluate whether the prediction model has practical value [29]. The decision curve was plotted using the ‘decision\_curve’ function in a R package of ‘rmda’, and the hyperparameters were set as follows: family = binomial (link = ‘logit’), thresholds = seq(0, 1, by = 0.01), study.design = ‘case-control’ and bootstraps = 100.

## 2.8. Statistical analysis

The normality and the homogeneity of variance between groups were tested for the continuous variables by using the ‘Shapiro’ function (Shapiro-Wilk normality test) and ‘Levene’ function, respectively. For the comparison between two independent samples, the student t-test was used for the continuous variables with normal distribution and homogeneous variance, otherwise, the Mann-Whitney *U* test was performed, and the counting data were analyzed with chi-square test. The ‘Pearson’ or ‘spearman’ functions were used to analyze the correlation between variables, and the ‘partial\_corr’ function was used for performing partial correlation analysis. When a *P* value was less than 0.05, the difference was considered statistically significant.



**Fig. 1.** Differential miRNAs were discovered with miRNA-TLDA assay and validated in a case-control study. (A) Heatmap of the plasma miRNAs in patients in pre-CAD-stage (PD1-E, PD2-E, and PD3-E) and CAD-stage (PD1-CAD, PD2-CAD, and PD3-CAD). (B) Volcanic map plotted with the mean relative folds change (x-axis) and the negative logarithm of *P*-value (y-axis). The points above the horizontal dotted line and to the left of the vertical dotted line represent significantly down-regulated miRNAs, and the points above the horizontal dotted line and to the right of the vertical dotted line represent significantly up-regulated miRNAs. (C) The statistical distribution of the relative plasma level of miR-140-3p in the CAD and control groups was presented with a box and whisker plots. The horizontal lines within the boxes represent the median value and the vertical lines extending below and above the boxes indicate 5–95% percentile values, respectively. \*\*\*, *P* < 0.001.

### 3. Results

#### 3.1. Significant association of miR-140-3p with the progress of CAD

In the present study, 182 of 380 miRNAs with a level higher than the detection limit were detected by TLDA in all six plasma samples. The plasma levels of the miRNAs were expressed with the mean multiples relative to that of miR-27b and miR-151a-3p. The expression profiles were an obvious difference between the plasma from the patients in pre-CAD stage and CAD stage (Fig. 1A). The plasma level of most miRNAs decreased, and less than 10% of miRNAs increased in CAD stage (Fig. 1B). Only 9 of the 182 miRNAs had a significant difference suggested by paired *t*-test ( $P < 0.05$ ) between the two groups, but which was no significance after Bonferroni correction ( $P > 0.05$ ). Hsa-miR-203 was significantly increased, and the others decreased obviously in CAD-stage (Table 1). Among the nine differentially expressed plasma miRNAs, miR-140-3p had a less standard error with a fold change  $>3$  and a higher statistical significance, thus it was selected for further investigation in a subsequent case-control study.

#### 3.2. Significantly negative association of the plasma miR-140-3p with CAD

A total of 205 samples, including 129 patients with CAD and 76 normal controls, were collected for validating the association of miR-140-3p with CAD. The clinical features and laboratory biochemical indicators included were shown in Table 2. The plasma level of miR-140-3p was expressed as the multiples relative to the plasma level of miR-151a-3p. The results show that the plasma level of miR-140-3p decreased significantly in the CAD group [1.26 (0.68, 2.01)] compared with the control group [2.07 (1.19, 3.21)] ( $P < 0.001$ ) (Fig. 1C and Table 2), which was consistent with the TLDA results. The plasma levels of miR-140-3p in the men sub-group and women sub-group were respectively [1.42 (0.80, 2.26)] and [1.61 (0.86, 2.51)], and were no significant difference ( $P > 0.05$ ), indicating the mismatch of gender would not affect the association between miR-140-3p and CAD. Pearson correlation analysis between variables shows that only the variable 'smoking history' was significantly correlated with miR-140-3p ( $r = -0.205$ ,  $P = 0.003$ ) (Table S2), this correlation was reproduced by Spearman's correlation analysis ( $\rho = -0.178$ ,  $P = 0.011$ ) (Table S3). However, significant correlations between HDL-C and variables such as gender, smoking history, age, TG, CHOL, APOA1, and APOB were observed (Tables S2 and S3). A significant correlation between miR-140-3p and CAD risk was shown by Pearson correlation analysis ( $r = -0.332$ ,  $P < 0.001$ ) (Table S2) and Spearman correlation analysis ( $\rho = -0.303$ ,  $P < 0.001$ ), respectively (Table S3). A partial correlation analysis by controlling all the differential variables such as gender, age, smoking history, TC, HDL-c, and LDL-c still showed a significant association between CAD risk and plasma level of miR-140-3p ( $r = -0.25$ , 95% CI [-0.37, -0.11],  $P < 0.001$ ). We also analyzed the associations between the plasma levels of miR-140-3p and clogged arteries, but no correlation was observed (data are not shown).

#### 3.3. miR-140-3p is independently associated with CAD

Binary logistic regression analysis showed that miR-140-3p, HDL-c, APOA1, and smoking history remained in the final equation and showed a statistical significance (Table 3), suggesting that they were independently associated with CAD. The accuracy of classification for the samples using binary logistic regression (BLR) with different variables was distinct (Fig. 2A). The AUCs of the BLR with miR-140-3p and HDL-c were 0.681 and 0.709, respectively. The AUCs of BLR with variable combinations such as HDL-c + APOA1, miR-140-3p + HDL-c + APOA1 and miR-140-3p + HDL-c + APOA1 + IsSmoking were 0.782, 0.831 and 0.860, respectively (Fig. 2A). The above results show that the accuracy of classification becomes higher when including more features in the BLR model, and the BLR model with miR-140-3p has similar accuracy for classification as that with HDL-c, suggesting miR-140-3p may be a novel marker for CAD risk as important as HDL-C.

LASSO regression analysis also showed that miR-140-3p, HDL-c, APOA1, and IsSmoking were required for the simplest model to classify the samples as CAD and control groups. miR-140-3p and HDL-c were the first two effective variables with nonzero entering the model (Fig. 2B). This indicates that miR-140-3p and HDL-C are the key features for classification or prediction for CAD.

**Table 1**

The differential plasma miRNAs between pre-CAD-stage and CAD-stage tested with miRNA-TLDA analysis.

Names	Pre-CAD-stage		CAD-stage		Folds change <sup>b</sup>	P
	Mean1 <sup>a</sup>	sd	Mean2 <sup>a</sup>	sd		
hsa-miR-140-3p	0.1052	0.0025	0.0131	0.0115	-3.01	0.0074
hsa-miR-26b-3p	0.0254	0.0078	0.0116	0.0085	-1.13	0.0400
hsa-miR-19b	51.6375	5.4834	19.6978	9.2391	-1.39	0.0475
hsa-miR-24	15.6452	6.1215	7.0213	2.9856	-1.16	0.0451
hsa-miR-126	3.2206	0.7447	1.6926	1.0859	-0.93	0.0253
hsa-miR-296	0.2162	0.1368	0.1212	0.1026	-0.83	0.0411
hsa-miR-376a	0.1366	0.0697	0.0774	0.0628	-0.82	0.0078
hsa-miR-139-5p	1.0375	0.2769	0.7986	0.2286	-0.38	0.0267
hsa-miR-203	0.0657	0.0695	0.1051	0.0709	0.68	0.0300

**Notes:** <sup>a</sup>, A mean  $\Delta$ Ct for each miRNA was calculated by averaging the  $\Delta$ Ct computed by the Ct of the candidate miRNAs minus the Ct of miR-151a-3p and miR-27b, respectively. Then the mean mean  $\Delta$ Ct was calculated among samples. <sup>b</sup>, The fold change was computed as the formula: Fold change (FC) =  $\log_2(\text{Mean2}/\text{Mean1})$ . A minus sign indicates a reduction.

**Table 2**  
Comparison of the difference in the variables between the CAD and the control groups.

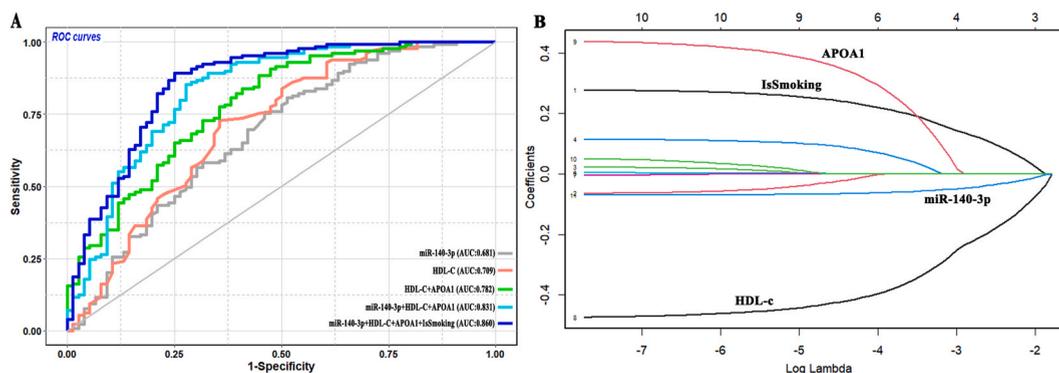
Variables	Control(n = 76)	CAD(n = 129)	P_values
Sex (Female %)	49 (64.5%)	43 (33.3%)	<0.001 <sup>a</sup>
Age (mean ± SD)	69.1 ± 9.2	65.8 ± 10.7	0.03 <sup>b</sup>
Hypertension	53(69.74%)	81(62.79%)	0.33 <sup>a</sup>
Diabetes	12 (15.79%)	28 (21.71%)	0.30 <sup>a</sup>
Hyperlipidemia	1 (1.32%)	3 (2.33%)	0.99 <sup>a</sup>
Smoking history	12(15.79%)	62(48.06%)	<0.001 <sup>a</sup>
Glucose (mg/dl) <sup>d</sup>	88.8(81.1, 102.9)	93.3(80.7, 113.3)	0.11 <sup>c</sup>
Triglyceride (mg/dl) <sup>d</sup>	116.0(88.6, 154.1)	129.3(89.5, 189.1)	0.12 <sup>c</sup>
Total cholesterol (mg/dl) <sup>d</sup>	191.03(164.0, 224.3)	164.9(145.2, 197.0)	<0.001 <sup>c</sup>
High density lipoprotein cholesterol (mg/dl) <sup>d</sup>	44.9(36.4, 54.1)	36.4(31.3, 42.2)	<0.001 <sup>c</sup>
Low density lipoprotein cholesterol (mg/dl) <sup>d</sup>	122.6(98.6, 143.9)	99.0(84.9, 127.4)	<0.001 <sup>c</sup>
APOA1 ( mg/dl ) <sup>d</sup>	120.0(110.0, 145.0)	119.0(107.8, 139.5)	0.36 <sup>c</sup>
APOB ( mg/dl ) <sup>d</sup>	89.0(76.0, 102.5)	84.5(73.0, 104.3)	0.29 <sup>c</sup>
miR-140-3p (folds relative to miR-151a-3p) <sup>d</sup>	2.07(1.19, 3.21)	1.26(0.68, 2.01)	<0.001 <sup>c</sup>

Note: a, chi-square test; b, T-test; c, Mannwhitney U test; d, quantile, median (25%, 75%).

**Table 3**  
Variables stained in the final equation with Logistic regression analysis.

	B	S.E	Walds	p-value	Exp (B)	95% C.I.	
						lower	upper
Smoking history	1.490	0.429	12.044	0.001	4.435	1.912	10.287
HDL-c	-7.398	1.332	30.859	<0.001	0.001	0.000	0.008
APOA1	6.556	1.343	23.829	<0.001	703.227	50.577	9777.751
miR-140-3p	-0.575	0.180	10.152	0.001	0.563	0.395	0.802
constant	0.741	0.932	0.631	0.427	2.098		

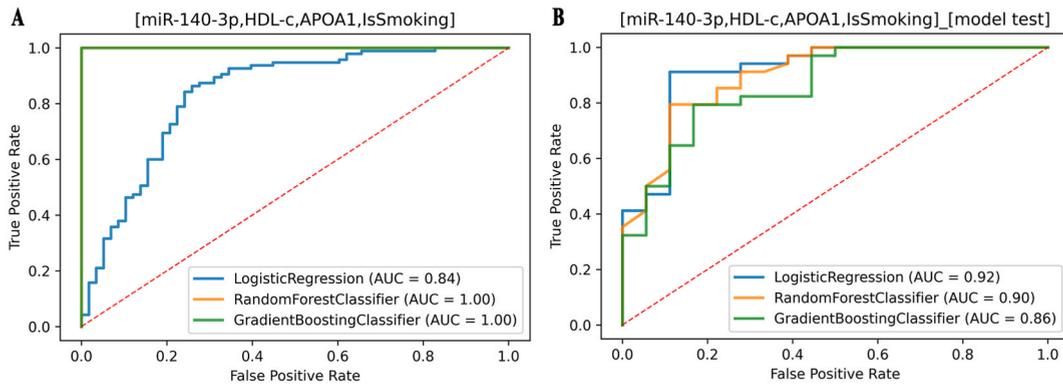
Note: B, regression coefficient; Exp(B), odds ratio; 95% C.I., 95% confidence interval of Exp(B).



**Fig. 2.** Importance of the variables independently associated with CAD in the classification models such as binary logistic regression and LASSO regression. (A) The ROC of prediction for the samples by binary logistic regression using different variables separately or in combination. (B) Important variables for classifying the samples in the simplest model were screened by LASSO regression.

3.4. The importance of miR-140-3p in prediction for CAD risk was validated by ML

In order to illustrate whether the four variables that were significantly associated with CAD have potential value in prediction for CAD risk, we constructed 3 ML models with LRC, RFC, and GBC using 'miR-140-3p', 'HDL-c', 'APOA1', and 'smoking history' in combination. The AUCs of the models were 0.84, 1.0, and 1.0, respectively (Fig. 3A), and the AUCs of the models tested with a test dataset were 0.92, 0.90, and 0.86, respectively (Fig. 3B). The importance of features determines their influence on model performance. Feature importance of 'miR-140-3p', 'HDL-c', 'APOA1', and 'smoking history' in RFC were 31.8%, 33.1%, 28.1%, and 7.0%, respectively (Fig. 4A), and that in GBC were 29.1%, 36.6%, 28.3%, and 6.0%, respectively (Fig. 4B). Obviously, the feature 'smoking history' is of the least importance in the models, and the importance of 'miR-140-3p' in the models is only less than that of 'HDL-c'. When excluding 'smoking history' from the features, using the rest three features to build an LRC, an AUC of 0.91 was obtained in the prediction for the test dataset (Fig. 5A). However, when excluding 'miR-140-3p' from the features, and using the rest three features to construct an LRC, an AUC was decreased to 0.73 in prediction for the test dataset (Fig. 5A). An AUC of 0.86 could be obtained when



**Fig. 3.** Evaluation of the performance of three ML models. (A) The ROC was plotted for LRC, RFC, and GBC in distinguishing CAD samples from the training dataset. (B) The ROC was plotted for LRC, RFC, and GBC in predicting CAD samples from the test dataset. [miR-140-3p, HDL-c, APOA1, IsSmoking] indicates the features used in combination to build the ML models. IsSmoking, smoking history.

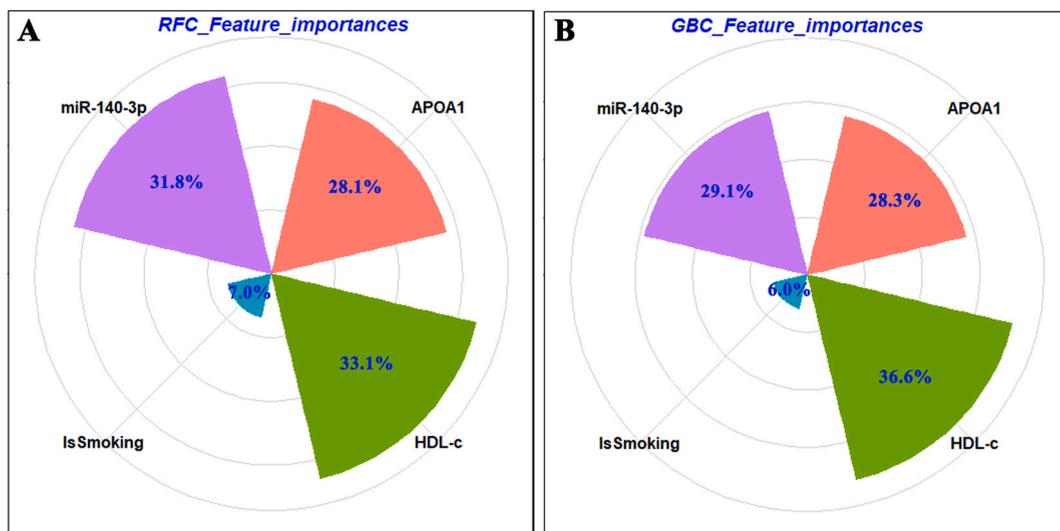
predicting with the test dataset using the LRC built with two features 'miR-140-3p' and 'HDL-c' (Fig. 5A). HDL-c is one of the TCRFs, and it was used to construct an LR model that has an AUC of 0.79 in prediction for the test dataset. An AUC of 0.77 was obtained when predicting the test dataset with an LRC built with 'miR-140-3p', this was only slightly lower than that of 'HDL-c' (Fig. 5B). All of these indicate that 'miR-140-3p' may be an effective indicator for CAD risk as important as 'HDL-C'.

### 3.5. Intervention considering the plasma level of mir-140-3p may acquire much net benefits

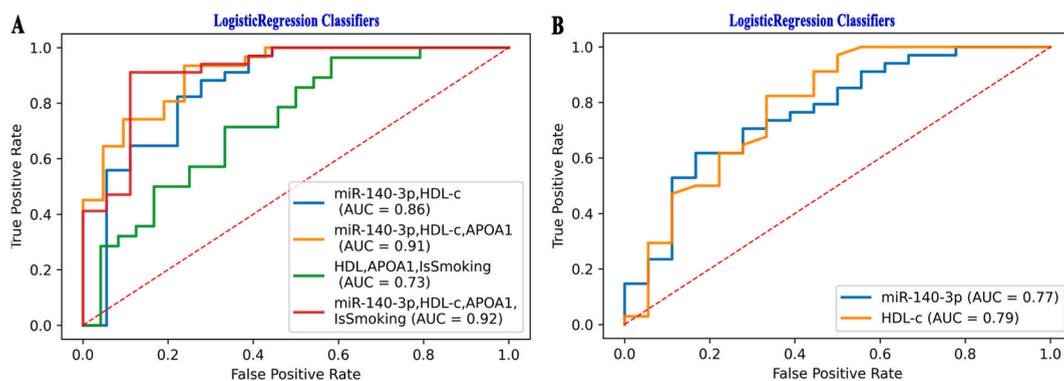
Decision curve analysis helps clinical decision makers balance the advantages and disadvantages of intervention and thus determine the best intervention point. It can also be used to evaluate whether a given model has practical value [29]. The decision curves for the 'logit' model constructed with 'miR-140-3p', 'HDL-c', 'HDL-c + APOA1', 'miR-140-3p + HDL-c + APOA1' and 'miR-140-3p + HDL-c + APOA1 + IsSmoking', respectively, were distant from the baseline for a wide range of risk thresholds (Fig. 6), suggesting that all of these models have high net benefits and potential practical value. The decision curves built with 'miR-140-3p' and 'HDL-c' almost overlap (Fig. 6), which again shows that the plasma level of miR-140-3p and HDL-C have similar practical value. The decision curves also show that the single feature, 'APOA1' or 'smoking history', has no practical value (Fig. 6).

## 4. Discussion

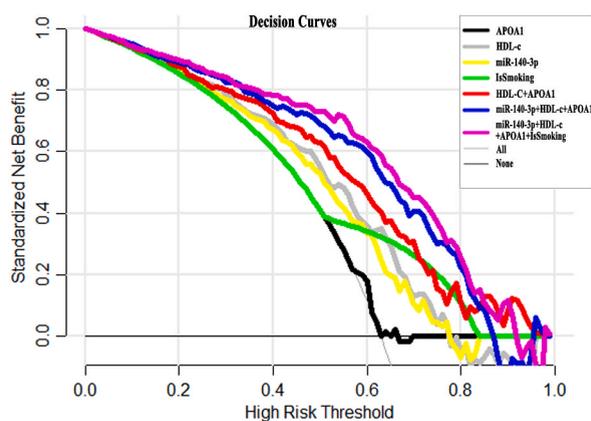
In the current study, decreased level of plasma miR-140-3p was found to be significantly associated with CAD. Moreover, among many factors, miR-140-3p and HDL-c were better indicators of risk stratification for individuals. miR-140-3p and HDL-c have similar



**Fig. 4.** Feature importance of the variables in RFC and GBC. (A) Feature importance of miR-140-3p, HDL-c, APOA1 and IsSmoking in random forest classifier; (B) Feature importance of miR-140-3p, HDL-c, APOA1 and IsSmoking in gradient boosting classifier. IsSmoking, smoking history.



**Fig. 5.** Comparison of performance of the logistic regression classifiers built with different variables separately or in combination. (A) Evaluating the influence of exclusion of different variables, respectively, on the performance of the classifier. (B) Comparison of performance of the classifiers built with miR-140-3p and HDL-c, respectively.



**Fig. 6.** Decision curves were constructed with different variables separately or in combination. The black horizontal line represents the net benefit varies with the high-risk threshold if none will be treated, and the gray curve represents the net benefit varies with the high-risk threshold if all will be treated. If the curve is distant from the gray curve and the horizontal lines, indicates the model will acquire net benefit in helping decision.

roles in the classification for the samples or in the prediction of CAD samples with different models, and which were used in combination in the models could significantly improve the accuracy compared to that they were used alone, respectively. Including miR-140-3p in the features was similar to including HDL-c in improving the performance of models such as LRC, RFC, GBC, and decision curve, and it was observed that miR-140-3p was much less affected by other variables than HDL-c. These strongly suggest that plasma miR-140-3p may be a potential novel biomarker associated with CAD risk as better than HDL-c.

Previous studies have shown that miR-140-3p has the functions of anti-oxidative stress, anti-inflammatory, and anti-apoptosis. Yi et al. [30] reported that over-expressing miR-140-3p significantly mitigated inflammation, oxidative stress, and cell apoptosis in an oxygen-glucose deprivation and reperfusion model. The correlation between decreased expression of miR-140-3p and osteoarthritis seems to have been fully proven. Ren et al. [31] discovered that miR-140-3p ameliorates the progress of osteoarthritis by targeting CXCR4. Peng et al. [32] demonstrated that amelioration of experimental autoimmune arthritis is through targeting synovial fibroblasts by intraarticular delivery of miR-140. Ntoumou et al. [33] found that miR-140-3p was significantly decreased in the serum of patients with osteoarthritis, and considered plasma miR-140-3p to be one of the potential biomarkers for the diagnosis of osteoarthritis. CAD is considered an inflammatory disease. In this study, it was found that the decrease in miR-140-3p plasma level was closely related to CAD. It is likely that the down-regulation of miR-140-3p led to the over-activation of a vascular endothelial inflammatory response, which promoted the occurrence and progression of CAD. In addition, it was also reported that miR-140-3p promotes the proliferation and migration of endothelial cells and the formation of endothelial tubes, mainly by directly inhibiting the expression of forkhead transcription factor FOXK2 [34]. miR-140-3p is reduced in In-Stent-Restenosis arteries and peripheral artery disease arteries, which results in the proliferation and apoptosis of smooth muscle cells by attenuating its inhibition on C-Myb and BCL-2 [35]. However, TAURINO et al. [36] reported significantly higher signal intensity of miR-140-3p in the whole blood of the CAD group than in that of healthy control through microarray analysis. Karakas et al. [37] observed that an increased serum level of miR-140-3p in patients with the acute coronary syndrome is associated with cardiovascular death. The harmful roles of miR-140-3p in these two studies seem not to

be supported by the functional roles, including anti-inflammatory, anti-oxidative stress, anti-apoptosis, and promoting angiogenesis, revealed by other studies. It is obvious that the previous studies on the role of miR-140-3p in CAD are controversial, that is mainly because there are differences in ethnicity, sample types, and disease stages in the study population among various studies. This study may be the first in the Chinese population and even in the East Asian population to report that the reduction of plasma miR-140-3p increases the risk of CAD.

A decrease in ApoA1 is considered the reason for the reduction of HDL-c and is also a risk factor for CAD [38]. In the present study, the binary logistic regression analysis showed an independent association between ApoA1 and CAD, but the models built with ApoA1 alone poorly distinguish CAD patients from the controls. This indicates that ApoA1 is not a good independent feature for CAD risk. In addition, we found that a low plasma level of miR-140-3p could also be used as an independent indicator of CAD risk like a low plasma level of HDL-c, a recognized CAD risk factor [39]. When the plasma HDL-c and miR-140-3p were used in combination to construct the ML model, the prediction accuracy of the model could reach 0.86. Plasma HDL-c level is one of the TCRFs and is affected by age and gender [39]. We found that in the classification of samples or in the prediction of unknown samples using a machine learning model, the use of indicator miR-140-3p alone has similar efficiency to that of HDL-C alone. Moreover, miR-140-3p and HDL-c were the first two features with non-zero coefficients entering the simplest models of LASSO regression, and the decision curves constructed with them almost overlap. These strongly suggest that the role of miR-140-3p in the prediction of CAD risk is not weaker than that of HDL-c. miR-140-3p was stable and existed in plasma with a high level, and its plasma level was not correlated with the other variables except for smoking history, suggesting that it may be an independent risk factor associated with CAD risk. The variability of miR-140-3p with age and gender also seemed to be less than that of HDL-C, thus it may be a better risk factor than HDL-c for risk prediction of CAD.

The plasma level of miR-140-3p was negatively associated with the smoking history that was one of the most definite risk factors for cardiovascular disease [40], this not only indirectly supported the roles of miR-140-3p in CAD risk but also indicated that smoking may induce the decrease of miR-140-3p. Smoking history is strongly associated with CVD risk [40], but it is not suitable as a general indicator of CVD risk because smokers are fewer and fewer in populations, especially in women [41]. miR-140-3p found to be closely related to CAD in this study is a genetic product in the human body and was not observed to change with age and gender, so it is expected to become a universal marker for predicting CAD risk.

In this study, we initially utilized paired samples whereby each participant was screened for differentially expressed miRNAs in their plasma before and after CAD diagnosis. We then validated our findings through case-control studies and employed various analytical methods, including ML, to determine the correlation between screened miRNAs and CAD, as well as their potential to predict CAD risk. All analytical methods yielded consistent and reliable results, giving us confidence in our conclusions.

## 5. Study limitations and conclusions

Limitations of this study may be that: Firstly, the miRNA-TLDA analysis could not detect all miRNAs, which caused many potential CAD-related miRNAs might be not detected, yet it's lucky that we discovered a CAD-related miRNA and validated its high prediction capacity in a case-control study; Secondly, this study is a single center study and the results could not be popularized before they are further validated in other studies with samples requiring being derived from a wide range of regions and ethnicity. Thirdly, there was an age and gender mismatch between the control and CAD groups. To account for this, we conducted a subgroup analysis by gender and found no significant difference in the plasma level of miR-140-3p between male and female subgroups. Furthermore, correlation analysis between plasma miR-140-3p levels and age indicated that age did not affect miR-140-3p plasma levels. Therefore, the results and conclusions are not impacted by the age and gender mismatch. In fact, it highlights the stability of plasma levels of miR-140-3p in comparison to HDL-c.

In conclusion, decreased level of plasma miR-140-3p is independently associated with CAD. Plasma miR-140-3p may be a novel biomarker of CAD risk as better than HDL-c, that can be used in combination with other CAD risk factors including HDL-c, APOA1, and smoking history to accurately predict CAD risk.

## Funding

This study was supported by [First-class Specialty Construction Project of Guangzhou Medical University], No. [02-408-2304-02079XM], and [Guangdong Provincial Science and Technology plan project], No. [2016A020215142].

## Data availability statement

The original data can be obtained from the corresponding author with reasonable request.

## Ethics statements

This retrospective study was reviewed and approved by the Institute Research Medical Ethics Committee of the Second Affiliated Hospital, Guangzhou Medical University (Approval No. 2016-ks-02) and conformed to the *principles outlined in the Declaration of Helsinki*. Written informed consent was waived for that this is a retrospective study.

## CRediT authorship contribution statement

**Pei Mo:** Writing – original draft, Data curation, Conceptualization. **Chao-Wei Tian:** Writing – original draft. **Qiqi Li:** Formal analysis, Data curation. **Mo Teng:** Data curation. **Lei Fang:** Investigation. **Yujuan Xiong:** Writing – review & editing, Funding acquisition, Conceptualization. **Benrong Liu:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yujuan Xiong reports financial support was provided by Guangdong Provincial Science and Technology. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

We are very grateful to Professor Yun Zhong and Professor Chengfeng Luo for their help in the collection of the clinical information of the patients.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e26960>.

## References

- [1] A.K. Malakar, D. Choudhury, B. Halder, P. Paul, A review on coronary artery disease, its risk factors, and therapeutics, *J. Cell. Physiol.* 234 (2019) 16812–16823.
- [2] S. Hosseinpour, B. Khalvati, F. Safari, A. Mirzaei, The association of plasma levels of miR-146a, miR-27a, miR-34a, and miR-149 with coronary artery disease, *Mol. Biol. Rep.* 49 (2022) 3559–3567.
- [3] K. Khalaj, R.L. Figueira, L. Antounians, G. Lauriti, Systematic review of extracellular vesicle-based treatments for lung injury: are EVs a potential therapy for COVID-19? *J. Extracell. Vesicles* 9 (2020) 1795365.
- [4] S. Kazemian, R. Ahmadi, G.A. Ferns, A. Rafiei, Correlation of miR-24-3p and miR-595 expression with CCL3, CCL4, IL1-beta, TNFalphaIP3, and NF-kappaB1alpha genes in PBMCs of patients with coronary artery disease, *EXCLI J.* 21 (2022) 1184–1195.
- [5] S. Kumar, M. Vijayan, J.S. Bhatti, P.H. Reddy, MicroRNAs as peripheral biomarkers in aging and age-related diseases, *Prog. Mol. Biol. Transl. Sci.* 146 (2017) 47–94.
- [6] T. Seeger, R.A. Boon, MicroRNAs in cardiovascular ageing, *J. Physiol.* 594 (2016) 2085–2094.
- [7] A. Kaur, S.T. Mackin, K. Schlosser, F.L. Wong, Systematic review of microRNA biomarkers in acute coronary syndrome and stable coronary artery disease, *Cardiovasc. Res.* 116 (2020) 1113–1124.
- [8] A. Rafiei, G.A. Ferns, R. Ahmadi, A. Khaledifar, Expression levels of miR-27a, miR-329, ABCA1, and ABCG1 genes in peripheral blood mononuclear cells and their correlation with serum levels of oxidative stress and hs-CRP in the patients with coronary artery disease, *IUBMB Life* 73 (2021) 223–237.
- [9] M. Hosseini, R. Sahebi, M. Aghasizadeh, D.F. Yazdi, Investigating the predictive value of microRNA21 as a biomarker in induced myocardial infarction animal model, *Gene Reports* 27 (2022).
- [10] S. Fichtlscherer, A.M. Zeiher, S. Dimmeler, Circulating microRNAs: biomarkers or mediators of cardiovascular diseases? *Arterioscler. Thromb. Vasc. Biol.* 31 (2011) 2383–2390.
- [11] C.J. Watson, S.K. Gupta, E. O'Connell, S. Thum, MicroRNA signatures differentiate preserved from reduced ejection fraction heart failure, *Eur. J. Heart Fail.* 17 (2015) 405–415.
- [12] C. Schulte, M. Karakas, T. Zeller, microRNAs in cardiovascular disease - clinical application, *Clin. Chem. Lab. Med.* 55 (2017) 687–704.
- [13] D.A. Velikiy, O.E. Gichkun, A.O. Shevchenko, MicroRNAs: a role in the development of cardiovascular disease, the possibility for clinical application, *Klin. Lab. Diagn.* 63 (2018) 403–409.
- [14] S.S. Zhou, J.P. Jin, J.Q. Wang, Z.G. Zhang, miRNAs in cardiovascular diseases: potential biomarkers, therapeutic targets and challenges, *Acta Pharmacol. Sin.* 39 (2018) 1073–1084.
- [15] S.P. Romaine, M. Tomaszewski, G. Condorelli, N.J. Samani, MicroRNAs in cardiovascular disease: an introduction for clinicians, *Heart* 101 (2015) 921–928.
- [16] M.W. Feinberg, K.J. Moore, MicroRNA regulation of atherosclerosis, *Circ. Res.* 118 (2016) 703–720.
- [17] M. Mohammad-Rezaei, R. Ahmadi, A. Rafiei, A. Khaledifar, Serum levels of IL-32 in patients with coronary artery disease and its relationship with the serum levels of IL-6 and TNF-alpha, *Mol. Biol. Rep.* 48 (2021) 4263–4271.
- [18] M. Valizadeh, M. Aghasizadeh, M. Nemati, M. Hashemi, The association between a Fatty Acid Binding Protein 1 (FABP1) gene polymorphism and serum lipid abnormalities in the MASHAD cohort study, *Prostaglandins Leukot. Essent. Fatty Acids* 172 (2021) 102324.
- [19] A. Rafiei, R. Ahmadi, S. Kazemian, T. Rahimzadeh-Fallah, Serum levels of IL-37 and correlation with inflammatory cytokines and clinical outcomes in patients with coronary artery disease, *J. Invest. Med.* 70 (2022) 1720–1727.
- [20] S. Kazemian, R. Ahmadi, A. Rafiei, F. Azadegan-Dehkordi, The serum levels of IL-36 in patients with coronary artery disease and their correlation with the serum levels of IL-32, IL-6, TNF-alpha, and oxidative stress, *Int. Arch. Allergy Immunol.* 183 (2022) 1137–1145.
- [21] G.S. Handelman, H.K. Kok, R.V. Chandra, A.H. Razavi, eDoctor: machine learning and the future of medicine, *J. Intern. Med.* 284 (2018) 603–619.
- [22] G. Quer, R. Arnaout, M. Henne, R. Arnaout, Machine learning and the future of cardiovascular care: JACC state-of-the-art review, *J. Am. Coll. Cardiol.* 77 (2021) 300–313.
- [23] A.M. Poss, J.A. Maschek, J.E. Cox, B.J. Hauner, Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease, *J. Clin. Invest.* 130 (2020) 1363–1376.
- [24] I. Shomorony, E.T. Cirulli, L. Huang, L.A. Napier, An unsupervised learning approach to identify novel signatures of health and disease from multimodal data, *Genome Med.* 12 (2020) 7.
- [25] B. Liu, L. Fang, Y. Xiong, Q. Du, A machine learning model based on genetic and traditional cardiovascular risk factors to predict premature coronary artery disease, *Front. Biosci.* 27 (2022) 211.

- [26] M. Saberi-Karimian, H. Safarian-Bana, E. Mohammadzadeh, T. Kazemi, A pilot study of the effects of crocin on high-density lipoprotein cholesterol uptake capacity in patients with metabolic syndrome: a randomized clinical trial, *Biofactors* 47 (2021) 1032–1041.
- [27] P. Mestdagh, T. Feys, N. Bernard, S. Guenther, High-throughput stem-loop RT-qPCR miRNA expression profiling using minute amounts of input RNA, *Nucleic Acids Res.* 36 (2008) e143.
- [28] H. Wan, Z. Zhao, C. Qian, Y. Sui, Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber, *Anal. Biochem.* 399 (2010) 257–261.
- [29] B. Van Calster, L. Wynants, J.F.M. Verbeek, J.Y. Verbakel, Reporting and interpreting decision curve analysis: a guide for Investigators, *Eur. Urol.* 74 (2018) 796–804.
- [30] M. Yi, Y. Li, D. Wang, Q. Zhang, KCNQ1OT1 exacerbates ischemia-reperfusion injury through targeted inhibition of miR-140-3p, *Inflammation* 43 (2020) 1832–1845.
- [31] T. Ren, P. Wei, Q. Song, Z. Ye, MiR-140-3p ameliorates the progression of osteoarthritis via targeting CXCR4, *Biol. Pharm. Bull.* 43 (2020) 810–816.
- [32] J.S. Peng, S.Y. Chen, C.L. Wu, H.E. Chong, Amelioration of experimental autoimmune arthritis through targeting of synovial fibroblasts by intraarticular delivery of MicroRNAs 140-3p and 140-5p, *Arthritis Rheumatol.* 68 (2016) 370–381.
- [33] E. Ntounou, M. Tzetis, M. Braoudaki, G. Lambrou, Serum microRNA array analysis identifies miR-140-3p, miR-33b-3p and miR-671-3p as potential osteoarthritis biomarkers involved in metabolic processes, *Clin. Epigenet.* 9 (2017) 127.
- [34] D. Wang, H. Wang, C. Liu, X. Mu, Hyperglycemia inhibition of endothelial miR-140-3p mediates angiogenic dysfunction in diabetes mellitus, *J. Diabet. Complicat.* 33 (2019) 374–382.
- [35] Z.R. Zhu, Q. He, W.B. Wu, G.Q. Chang, MiR-140-3p is involved in in-stent restenosis by targeting C-myc and BCL-2 in peripheral artery disease, *J. Atherosclerosis Thromb.* 25 (2018) 1168–1181.
- [36] C. Taurino, W.H. Miller, M.W. McBride, J.D. McClure, Gene expression profiling in whole blood of patients with coronary artery disease, *Clin. Sci. (Lond.)* 119 (2010) 335–343.
- [37] M. Karakas, C. Schulte, S. Appelbaum, F. Ojeda, Circulating microRNAs strongly predict cardiovascular death in patients with coronary artery disease-results from the large AtheroGene study, *Eur. Heart J.* 38 (2017) 516–523.
- [38] E.J. Rhee, C.D. Byrne, K.C. Sung, The HDL cholesterol/apolipoprotein A-I ratio: an indicator of cardiovascular disease, *Curr. Opin. Endocrinol. Diabetes Obes.* 24 (2017) 148–153.
- [39] M.J. Legato, Dyslipidemia, gender, and the role of high-density lipoprotein cholesterol: implications for therapy, *Am. J. Cardiol.* 86 (2000) 15L–18L.
- [40] K.K. Teo, T. Rafiq, Cardiovascular risk factors and prevention: a perspective from developing countries, *Can. J. Cardiol.* 37 (2021) 733–743.
- [41] J. He, Z. Zhu, J.D. Bundy, K.S. Dorans, Trends in cardiovascular risk factors in US adults by race and ethnicity and socioeconomic status, 1999–2018, *JAMA* 326 (2021) 1286–1298.