scientific reports

Check for updates

OPEN

Integration of short non coding RNA and genetic factors for coronary artery disease risk prediction in a prospective study

Elisabetta Casalone[1]✉, Miriam Rosselli[1], Giovanni Birolo[2], Carla Debernardi[1], Chiara Catalano[1], Serena Aneli[3], Alessandra Allione[1], Cecilia Di Primio[1], Annette Peters[4,5,6], Christian Gieger[4,7], Gabrielle Anton[4,7], Paolo Vineis[8], Carlotta Sacerdote[9] & Giuseppe Matullo[1,10]✉

Coronary artery diseases (CADs) continue to be the leading global contributors to multi-morbidity and mortality. Given the significant burden of CADs, there is a critical need to identify novel and effective biomarkers for risk assessment. This study sought to evaluate the potential of serum extracellular vesicle-derived small non-coding RNAs (sncRNAs) as predictive biomarkers for CAD risk. Using next-generation sequencing approach, the levels of extracellular vesicles (EVs)-associated sncRNAs were analysed in serum samples from 91 pre-clinical CAD cases and their matched healthy controls, sourced from the prospective EPICOR cohort. We evaluated the predictive ability of sncRNAs alone and in combination with polygenic risk score (PRS) PGS000329. We identified 44 differentially expressed microRNAs (miRNAs) and PIWI-interacting RNAs (piRNAs) (FDR < 0.05), which were then narrowed down to ten significant signals (|log2FC|>0.6) for technical validation. RT-qPCR analysis confirmed the trend of expression for two miRNAs (miR-194-5p and miR-451a) and six piRNAs (piR-20266, piR-23533, piR-27282, piR-28212, piR-1043, piR-619). The ROC curve from a Random Forest model showed a higher discrimination ability of piR-619 and piR-23,533 (AUC = 0.72) compared to the use of traditional risk factors alone (AUC = 0.68). To enhance CAD risk assessment, we integrated genetic data by stratifying the cohort into two groups based on the 80th percentile of the PGS000329. We observed an odds ratio (OR) of 2.8 (95% CI: 1.3–6.4, $p = 0.01$) using PGS000329 alone. When the model was adjusted to include two piRNAs and smoking status, the OR increased to 3.26 (95% CI: 1.2–9.5, $p = 0.02$). Even though this study is limited by the absence of an independent replication cohort, these findings suggest that the two piRNAs pattern could contribute to predict the risk of CAD and may provide valuable insights into the underlying pathogenesis of the disease, in particular integrating individual CAD-PRS.

Keywords Coronary artery disease, PiRNAs, MiRNAs, Polygenic risk score, Biomarkers, NGS

Cardiovascular diseases (CVDs) are a leading cause of mortality, morbidity, and hospitalization in the adult population in Western countries and a major challenge for low- and middle-income countries that follow a Westernised-lifestyle[1].

[1]Genomic Variation, Complex Diseases and Population Medicine, Department of Medical Sciences, University of Turin, Turin, Italy. [2]Computational Biomedicine, Department of Medical Sciences, University of Turin, Turin, Italy. [3]Department of Public Health Sciences and Pediatrics, University of Turin, Turin, Italy. [4]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Munich, Germany. [5]Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany. [6]Munich Heart Alliance, German Center for Cardiovascular Disease (DZHK E.V.), Partner-Site Munich, Munich, Germany. [7]Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg, Munich, Germany. [8]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [9]Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin, Italy. [10]Medical Genetic Service, Città Della Salute E Della Scienza, Turin, Italy. ✉email: elisabetta.casalone@unito.it; giuseppe.matullo@unito.it

Coronary artery diseases (CADs) are the most common forms of CVD with a complex and multifactorial etiology, which includes numerous environmental risk factors, as well as a strong genetic background[2].

Particularly, recent studies have shown that Polygenic Risk Scores (PRS) improve CAD prediction and identification of high-risk individuals[3]. It remains of interest to evaluate how the combined PRS with additional biomarkers can improve the predictive accuracy of the established CAD PRS.

There has been increasing evidence that circulating small non coding RNAs (sncRNAs) can be used as diagnostic and prognostic biomarkers for several diseases, including CADs. Among these molecules, the class of microRNAs (miRNAs) is the most studied, and numerous works reported their involvement as biomarkers in CVDs[4,5]. Since the finding that miRNAs are present in the circulation, they have been investigated as novel biomarkers, especially in the context of acute myocardial infarction and heart failure.

Although there is limited compelling evidence to suggest that miRNAs can surpass traditional biomarkers in terms of predictive power, there is considerable potential for miRNAs, when used in conjunction with newly developed PRS, to enhance existing risk prediction models and improve the accuracy of individual risk assessment[6,7].

The opportunity to use miRNAs as biomarkers lies in their potential to discriminate myocardial infarction from other diseases, which have similar classical biomarker profiles. A number of miRNAs have been explored in myocardial infarction patients, but they were not as specific as compared to troponin[8]. In contrast, Devaux et al. found that miR-208b and miR-499a have a higher accuracy in discriminating myocardial infarction patients from those with other types of acute chest pain compared to troponins and they appear to be very early markers of cardiac damage[9]. Therefore, it seems that miRNAs still have some additional value on top of troponins that requires further investigation.

Nonetheless, it is particularly complex to draw a firm conclusion for the feasibility of miRNAs in clinical settings, due to the lack of standardization methods in analytical workflows. The most commonly used technique for measuring circulating miRNAs is qPCR, with its inherent limitations raised by the lack of unequivocally accepted normalization strategies[10].

The sncRNAs family also includes Piwi-Interacting RNAs (piRNAs), a class of small molecules which interact with the PIWI proteins to form a piRNA-induced RNA silencing complex[11]. Recent findings suggest that piRNAs may regulate ischemic heart disease acting upon the downstream targets of the AKT pathway, leading to a rise in studies regarding piRNAs in this disease[12]. Moreover, the altered expression of piRNAs (piR – 9010, piR – 28646, and piR- 23619) in the serum of heart failure patients suggests a critical regulatory function for these molecules and highlights their potential as biomarkers in cardiovascular diseases, in particular an increase expression of piR-2,106,027 was associated with the release of cardiac troponin I, indicating myocardial injury[13–15]. Most of circulating sncRNAs are released in bloodstream packaged in extracellular vesicles (EVs), small lipid bilayer membranes released by all cell types, with a relevant role in cell to cell communication, particularly during pathological conditions, promoting inflammation and angiogenesis[16]. EV-derived sncRNAs present higher specificity and stability than circulating molecules, they are protected from RNases degradation and reflect the pathological state of the cells and tissues of origin. Thus, these characteristics make sncRNAs carried by EVs ideal biomarkers for disease diagnosis.

In the present study we aimed at investigating the CAD predictive potential of sncRNAs detected in circulating EVs derived from serum samples of subjects belonging to the EPICOR prospective cohort. To understand whether these sncRNAs could be used as potential early biomarkers for CAD, we analysed the expression levels of miRNAs and piRNAs extracted from serum EVs by means of deep sequencing, comparing pre-clinical individuals who later developed CVD with matched healthy controls.

Moreover, we also evaluated the combined power of sncRNAs and PRS to better stratify the individual risk for prevention and early intervention.

Results

Baseline demographic and clinical characteristics of study cohort

The study cohort included 182 paired subjects, grouped into two clinical groups (91 pre-clinical cases and 91 matched healthy controls) belonging to the EPICOR - Turin cohort. Baseline relevant characteristics across the study population are summarized in Table 1.

Most of the 182 paired individuals were males (84.62%) and the majority of cases were smokers (53.85%). Among cases, 54.95% were heavy drinkers, 36.26% were light drinkers, while only 8.79% were classified as non-drinkers.

The mean age at recruitment was 50.5 years. Among the 91 cases, 32 suffered from Acute Myocardial Infarction (AMI), 9 endured Cardiac Revascularization, while 50 showed both events. Clinical variables such as BMI, blood pressure, total cholesterol, triglycerides and LDL levels were higher in pre-clinical cases of CAD compared to controls, whereas HDL levels appear to be higher in the control group.

The concentration and integrity of EVs after Exoquick precipitation were assessed in representative samples by NanoSight and TEM (tary Fig. S1). Quantitative size analysis showed that EVs have a mean of 40 nm and TEM analysis confirmed the typical cup-shaped morphology of EVs (Supplementary Fig. S2).

Next generation sequencing generated an average of 16 million reads per sample, ranging from 2.7 to 60.9 million reads. After adapter trimming, reads shorter than 14 bases were discarded, with an average of 60% retained clean reads per sample. Clean reads were then aligned and assigned to a unique miRNA or piRNA to produce the count matrix. The 2% of clean reads were successfully assigned on average per sample (Supplementary Table S1). Of these, 60% were assigned to miRNAs and the remaining 40% to piRNAs.

Overall, NGS analysis led to the identification of 172 sncRNAs, of which 132 were miRNAs and 40 were piRNAs. Among these, 44 were significantly differentially expressed after adjustment for the CAD-related

[*N* (%)]	Pre-Clinical CAD 91(50%)	Controls 91(50%)	*p*-value
Sex [N (%)]			
Female	14(15.38%)	14(15.38%)	
Male	77(84.62%)	77(84.62%)	
Smoke [N (%)]			
Ex-Smoker	26(28.57%)	35(38.46%)	
Non-Smoker	16(17.58%)	28(30.77%)	
Smoker	49(53.85%)	28(30.77%)	
ALCOHOL [N(%)]			
Non-Drinker (alcohol= 0 g/die)	8 (8.79%)	2 (2.2%)	
Light Drinker (0 < alcohol < 12 g/die)	33 (36.26%)	35 (38.46%)	
Heavy Drinker (alcohol > 12 g/die)	50 (54.95%)	54 (59.34%)	
AGE [years, mean ± sd]	50.88+/-6.63	50.04+/-6.46	
Pad [mean ± sd] (mmHg)	85.45+/-9.29	83.77+/-10.02	
Pas* [mean ± sd] (mmHg)	138.49+/-16.89	133.90+/-17.25	0.04
BMI* [mean ± sd] (kg/m2)	27.08+/-3.44	26.10+/-3.74	0.04
Total Cholesterol* [mean ± sd] (mmol/L)	237.55+/-43.77	219.27+/-37.70	0.02
Triglycerides* [mean ± sd] (mmol/L)	153.72+/-81.41	137.75+/-89.95	0.05
HDL* [mean ± sd] (mmol/L)	52.76+/-10.84	59.47+/-18.13	0.04
LDL* [mean ± sd] (mmol/L)	154.04+/-37.45	132.25+/-34.52	0.0004
Waist hip ratio	0.92+/-0.08	0.90 +/-0.08	
Events developed at follow-up [N (%)]			
Acute Myocardial Infarction	32 (35%)		
Cardiac Revascularization	9 (10%)		
Acute Myocardial Infarction and Cardiac Revascularization	50 (55%)		

Table 1. Clinical characteristics of the cohort. Pad: Diastolic blood pressure; Pas: Systolic blood pressure, BMI: Body Mass Index; HDL: High-Density Lipoproteins; LDL: Low-Density Lipoproteins; AMI: Acute Myocardial Infarction. *Differences between the two groups of parameters were statistically significant with p-value < 0.05.

variables and batch effect (FDR < 0.05) (Fig. 1, Supplementary Table S2), in particular 31 miRNAs were upregulated in pre-clinical CAD cases compared to healthy controls, while 14 piRNAs were downregulated.

A panel of four miRNAs (hsa-miR-451a, hsa-miR-375-3p, hsa-miR-194-5p and hsa-miR-22-5p, Supplementary Fig. S3a) and six piRNAs (hsa-piR-20266, hsa-piR-23533, hsa-piR-27282, hsa-piR-28212, hsa-piR-1043 and hsa-piR-619, Supplementary Fig. S3b) was tested by RT-qPCR in 20 pre-clinical cases and their 20 matched controls, randomly selected from a subset of 182 samples. The 10 sncRNAs were chosen according to |log2FC|>0.6 and FDR < 0.05 criteria.

The expression for miR-22-5p and miR-375-3p was not detectable by RT-qPCR. For eight signature sncRNAs, correlation analysis showed a good concordance (Rho ≤ -0.65) between log-transformed read counts of NGS and ΔCt values of RT-qPCR expression levels (Supplementary Table S3). Moreover, the trend of expression was comparable to those obtained by NGS (Fig. 2); however, the adjusted GLM reported a significant lower expression only for piR-619, piR-1043, piR-23,533, piR-27,282, piR-28,212 in pre-clinical CADs compared to controls (FDR < 0.05, Fig. 2).

Moreover, the five significantly differentially expressed piRNAs were highly correlated to each other (Supplementary Fig. S4). For further analysis we took into account piR-619, which was the most significant in GLM (Fig. 2) and piR-23,533, which showed a weaker correlation with the other four piRNAs.

MiRNAs target genes and enrichment analysis

Functional enrichment analysis was performed to explore if miR-194-5p and miR-451a are involved in relevant molecular networks related to the pathogenesis of CADs.

Thirty-nine target genes were validated by at least two out of three databases (miRecords[17], miRTarBase[17], and TarBase[18], Supplementary Table S4) and were further investigated through ShinyGO v0.80 and WEB-based GEne SeT AnaLysis Toolkit (WebGestalt) online tools using The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology Biological Process (GO-BP) databases[19].

The analysis of target genes in GO-BP revealed the identified target genes were involved in several key pathways, which were narrowed down to five main categories using Weighted Set Cover Redundancy Reduction method. Significantly enriched (FDR ≤ 0.05) pathways are reported in Fig. 3, along with target genes involved in the pathway and relative FDR. The full results of significantly enriched GO-BP from ShinyGO v0.80 are reported in Supplementary Table S5.

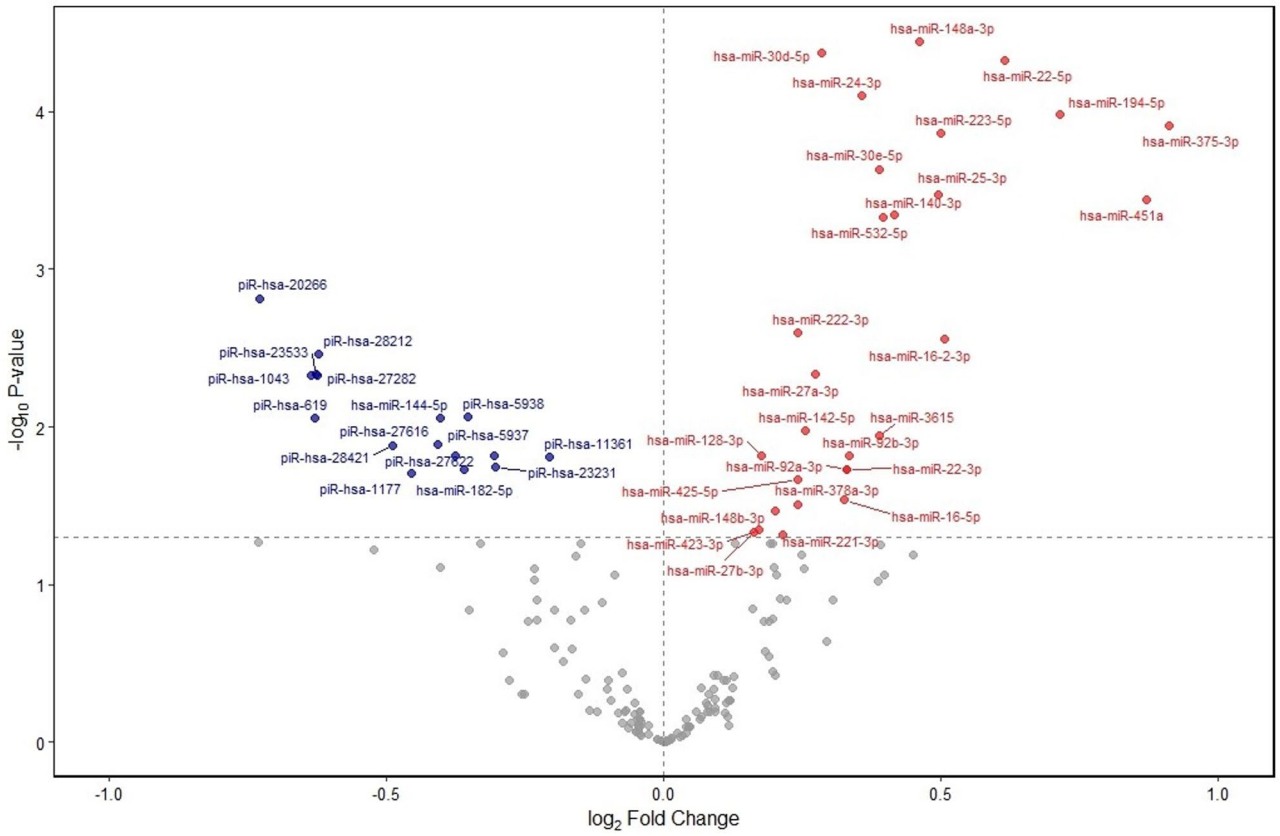

Fig. 1. Volcano plot representing the differentially expressed sncRNAs. Colored dots represent the 44 statistically significantly deregulated sncRNAs (FDR < 0.05). The upregulated sncRNAsare represented in red, the downregulated in blue. The gray dots represent sncRNA that were not significantly differentially expressed at FDR.

Similarly, the KEGG pathway analysis indicated that target genes are significantly involved in several key processes, such as "JAK-STAT, FoxO and PI3K-Akt signalling pathways" (Fig. 4). The full results of significantly enriched KEGG from ShinyGO v0.80 are reported in Supplementary Table S6.

Prediction model

The discriminative capability of the validated piRNAs in distinguishing preclinical CAD cases from healthy controls was tested with a Random Forest classification model. The model was trained in cross-validation using the same samples of the discovery phase due to the unavailability of a replication cohort, including patient age, sex, and anthropometric variables alone or in combination with piR-619 and piR-23,533. The combination of the data provided a better classification performance (AUC = 0.72 ± 0.01) compared to the clinical variables alone (AUC = 0.68 ± 0.01) (Fig. 5).

Integration of PRS with PiRNAs expression profile

We estimated the CAD risk in the EPICOR Turin cohort using the PGS000329. The genetic data were available for 177 individuals (88 pre-clinical CAD, 89 controls). The score distributions between the two groups were significantly different (Kolmogorov–Smirnov test p-value = 0.007, Supplementary Fig. S5). The mean score for the control group was 0.86, while for cases it was 1.40 (p-value < 0.0001); the mean score was 1.40 and 0.86 for pre-clinical CAD and control group, respectively.

We split the data above and below the 80th percentile and the individuals above this threshold showed an OR of 2.8 (95%CI 1.3–6.4).

Incorporating piR-619 and piR-23,533 into the model to assess their potential impact on risk estimation yielded an odds ratio (OR) of 3.26 (95% CI: 1.2–9.5). When additional variables such as age, sex, waist-to-hip ratio, alcohol consumption, smoking status, hypertension, treatment for hyperlipidaemia, and total cholesterol were included in the model, only smoking status demonstrated a significant effect ($p < 0.0001$). Nevertheless, the influence of piR-619 ($p = 0.06$) and piR-23,533 ($p = 0.0008$) persisted.

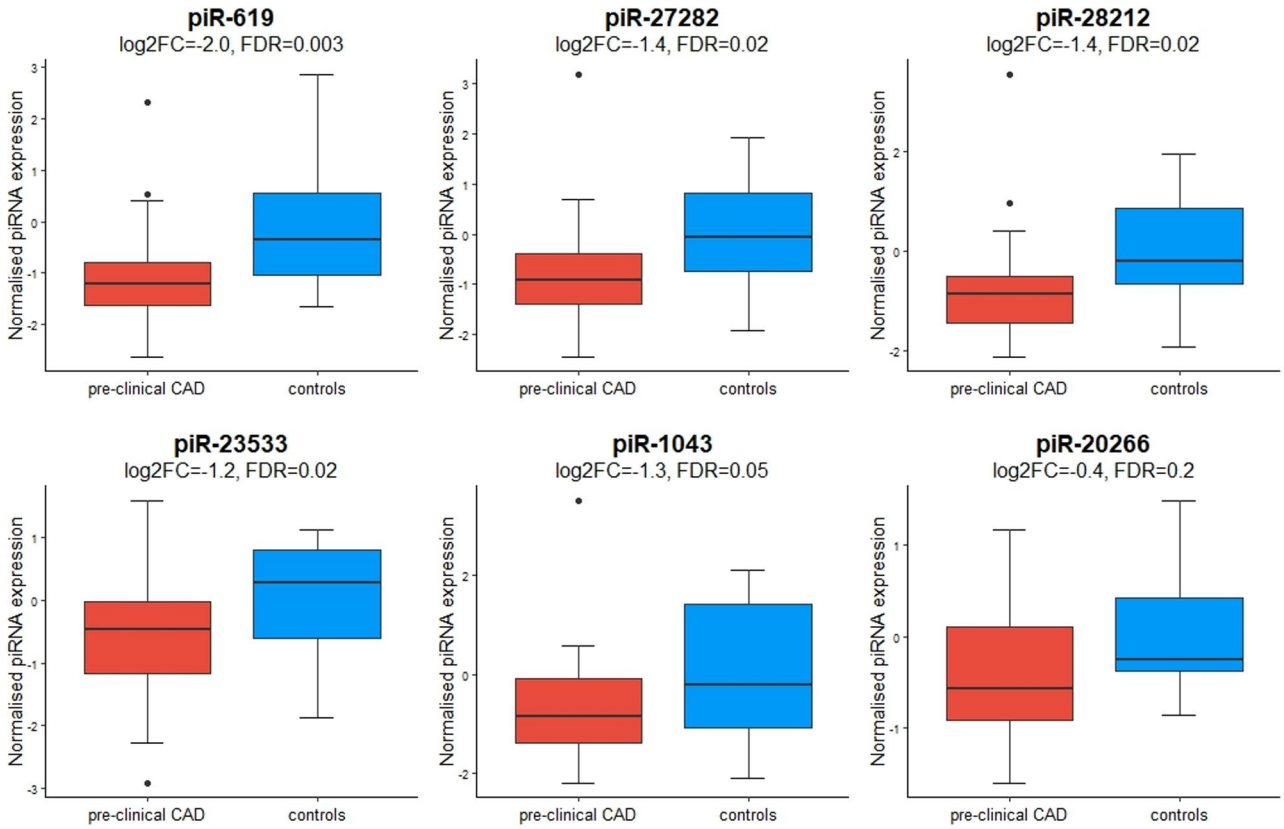

Fig. 2. Boxplots of the significantly differentially expressed piRNAs (FDR < 0.05) obtained in RT-qPCR.

Discussion

The objective of this study was to identify potential predictive biomarkers among serum EV-derived sncRNAs to predict CAD risk. Currently, there are few biomarkers which may lead to the identification of individuals who are at risk of CAD were identified, however the traditional risk factors such as age, cholesterol levels, blood pressure, smoking, and diabetes, may not capture the full spectrum of risk determinants, failing to predict CAD events in a substantial number of cases. The role of sncRNAs especially miRNAs has been investigated in several studies but most of them have focused on serum or plasma cell-free miRNAs in patients after the coronary event for diagnostic or prognostic purposes[20]. Moreover, the absence of detailed clinical information precluded the determination of a predictive value of the suggested miRNAs.

We performed a large-scale sncRNAs profiling by deep sequencing of 91 pre-clinical CAD cases and 91 matched non-disease controls belonging to the prospective EPICOR Turin cohort, with detailed measures of clinical parameters involved in the risk of CAD. Overall, we identified 44 miRNAs and piRNAs significantly differentially expressed between the pre-clinical cases compared to healthy controls. RT-qPCR analysis confirmed a statistically significant lower expression of piR-1043, piR-23,533, piR-28,212, piR-27,282, piR-619 in pre-clinical cases, while showed the same expression trend for miR-194-5p, miR-451a, piR-20,266. We identified piR-619 and piR-23,533 as the minimal set of sncRNAs able to identify individuals at risk of developing CADs.

The advent of high-throughput technologies in recent decades has significantly advanced our understanding of noncoding genomes[21]. A critical aspect of this research is the piRNA/PIWI pathway, which plays a pivotal role in regulating the homeostasis of mRNA, lncRNA, and satellite RNA through interactions with transposons and pseudogenes via both transcriptional and posttranscriptional mechanisms.

Investigating the diverse functions of piRNAs in the pathogenesis of CAD and other cardiovascular diseases holds significant potential for clinical application. PIWI proteins and piRNAs are expressed in cardiomyocytes and exhibit dynamic regulation in cardiovascular conditions, with piRNAs also implicated in the regulation of DNA methylation in heart disease.

Recent evidence indicates that piRNAs are involved in cardiac fibrosis and hypertrophy, and may regulate myocardial ischemia by targeting LINE-1 in cardiac progenitor cells with an activation of AKT pathway and inhibition of cell apoptosis[22].

Differences in serum exosomal piRNA profiles between patients with heart failure and healthy individuals further highlight their critical roles in disease onset and progression[23]. Notably, four miRNAs were upregulated in the serum of myocardial infarction patients compared with healthy controls, but not in individuals with acute heart failure or stable coronary artery disease, suggesting a disease-specific proinflammatory piRNA signature with potential as biomarkers and therapeutic targets. Moreover, the combination of multiple sncRNAs, including piRNAs, has demonstrated promise as a prognostic biomarker across different stages of heart failure[12].

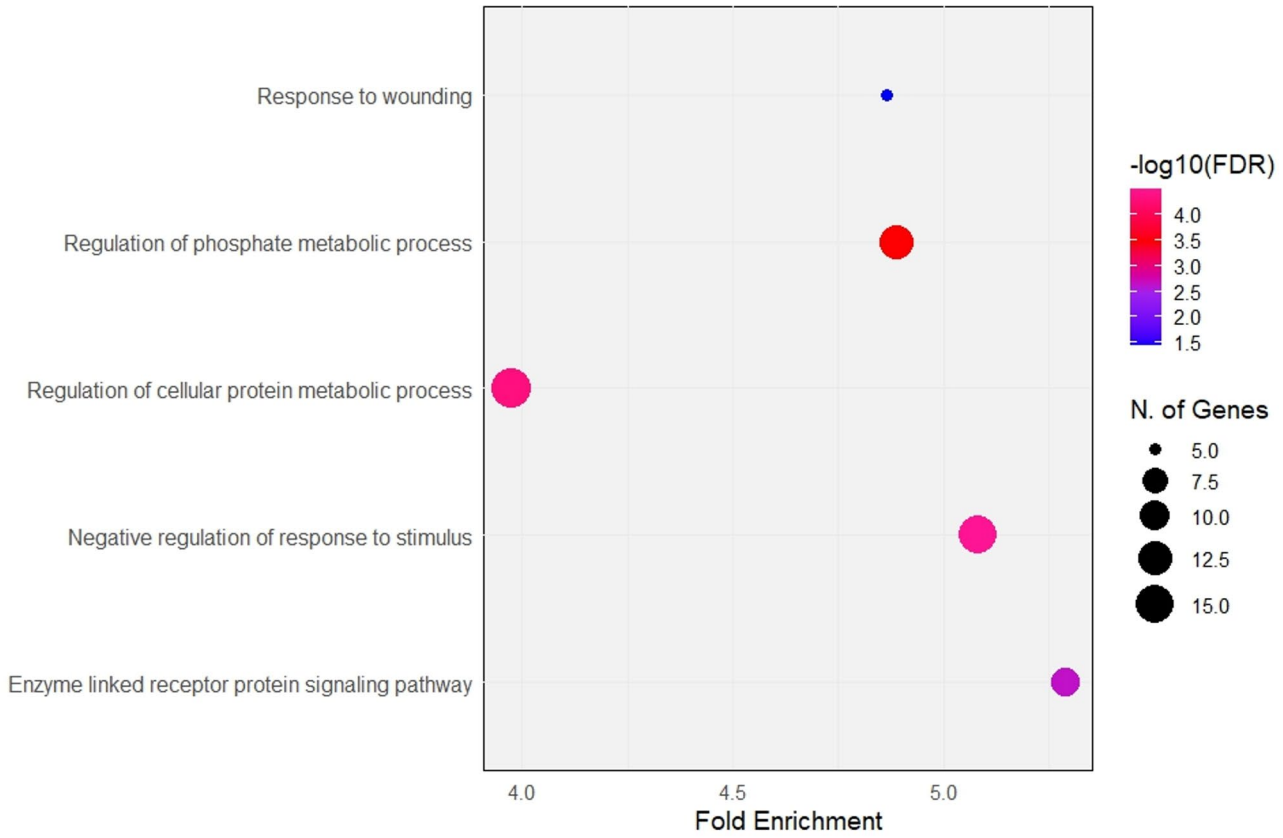

Fig. 3. Enrichment analysis of Gene Ontology Biological Process (GO-BP) pathways relative to the validated target genes of miR-194-5p and miR-451a. The figure shows the most relevant GO-BP pathways, along with the relative involved target genes and FDR. The size of the circle is proportional to the number of enriched genes.

In a study focused on head and neck carcinoma, the authors reported that piR-23,533 directly interacts with USP7, regulating cell proliferation and apoptosis[24]. USP7 is a deubiquitinating enzyme that plays a pivotal role in myocardial injury, particularly within cardiomyocytes, where it promotes fibrosis through a Smad3-independent KLF7–GATA3 signaling axis[25]. Given the established involvement of USP7 in human fibrotic myocardium and its putative regulation by piR-23,533, may provide a mechanistic link between this piRNA and pathways relevant to coronary artery disease.

Although these findings remain preliminary, they support the plausibility that piR-23,533 could influence CAD risk through modulation of USP7-related molecular networks and further experimental validation will be required to confirm this interaction and to elucidate its downstream functional consequences in cardiovascular tissues.

So far, most of the studies have focused on serum proteins and plasma cell-free miRNAs in patients after the coronary event for diagnostic or prognostic purposes[26,27], but the absence of detailed clinical information precluded the determination of a predictive value linked to the suggested miRNAs.

We identified miR-451a and miR-194-5p by NGS as differentially expressed in pre-clinical CAD compared with controls. The role of miR-451a and miR-194-5p in CVD was already described in several works. Taraldsen et al. found miR-451a to be related to heart plaque formation and progression[28]. Moreover, Marques and colleagues showed that miR-451a had high specificity and sensitivity in the clinical diagnosis of heart failure[29]. Mir-194-5p has been described by Neiburga et al. [30] as significantly correlated to CAD risk factors and Streese et al. associated high expression of miR-194-5p with a higher CAD risk profile[26]. In vitro experiments demonstrated that miR-194 transfection inhibited apoptosis and restored the cell cycle, indeed among the target genes there are key regulators of cell cycle and apoptosis[31].

Functional enrichment analysis of the validated miRNA target genes reported a different set of pathways (ErbB2, PI3K-AKT, JAK-STAT and mTOR signalling, pathways regulating pluripotency of stem cells such as WNT signalling pathway, focal adhesion pathways, endocrine resistance and endocytosis) and processes (regulation of response to stimuli, of protein metabolic processes, response to wounding and enzyme-linked receptor protein signalling pathways) related to CAD. The JAK/STAT signalling pathway plays a regulatory function in myocardial ischemia-reperfusion injury, fibrosis and heart failure[32,33].

Several studies suggest that the WNT pathway plays a major role in the progression of heart disease, both in terms of both metabolic alterations (insulin sensitivity) and cardiovascular remodelling as well as structural changes (fibrosis, sclerosis, atheroma formation, smooth muscle cell proliferation, hypertrophy)[34]. Increased

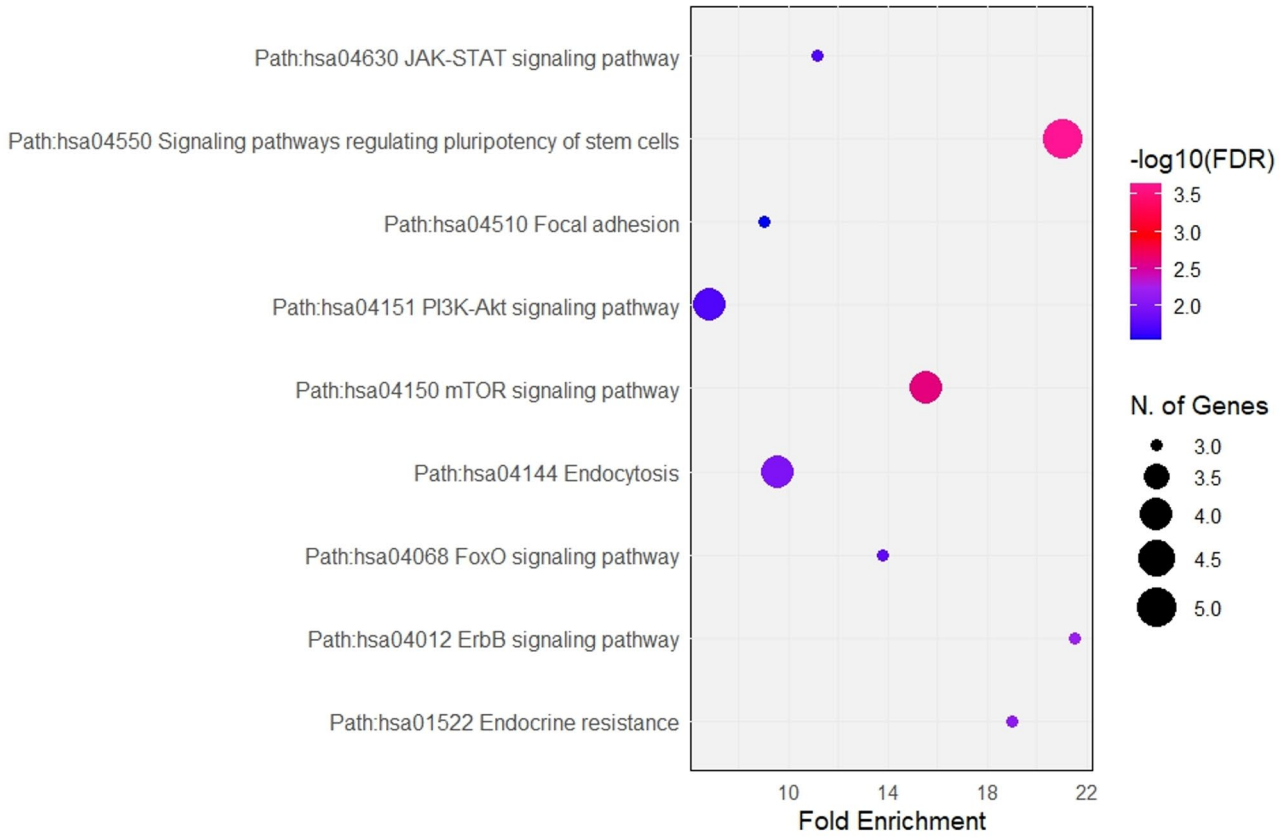

Fig. 4. KEGG pathway analysis of validated miRNAs target genes. The figure shows the most relevant KEGG pathways[46–48], along with the relative involved target genes and FDR. The size of the circle is proportional to the number of enriched genes.

serum HER2/ErbB2 levels appear to be strongly associated with the presence of CAD, particularly in people with obesity, while others suggest that mTORC1 regulates cardiac adaptation to energy deprivation and ischaemia, contributing to pathological cardiac remodelling in response to chronic ischaemic injury[35,36].

These results strengthen our hypothesis, demonstrating that target genes, thus the sncRNAs linked to them, are indeed key pathways in CAD development.

As previous works suggested, single sncRNAs are oftentimes insufficient in determining disease risk. Consistently, the detected piRNAs in this context exhibit limited discriminative capacity and there is a need for large-scale studies to validate piRNAs as reliable biomarkers across different populations and conditions.

Moreover, accurate risk prediction requires consideration of additional factors, including the individual's genetic background, which can significantly influence disease susceptibility and biomarker expression. PRSs have proven to be useful for capturing the individual's genetic susceptibility to CAD development, and it remains of high interest to determine the potential of integrating disease-related molecules with specific PRS to enhance disease prediction performance.

In the present study, we assessed the potential of risk prediction by PRS[37] with piRNA expression profiles.

This integrative approach yielded an OR of 3.26, indicating a substantially elevated risk in individuals with combined genetic and molecular risk factors compared to those without. Although the sample size was limited, our findings indicate that combining PRS with piRNAs profiles and clinical data may significantly enhance risk stratification compared to relying on genetic information alone.

The integrated PRS–piRNA approach could enable a more comprehensive characterization of individual risk, enhancing patient stratification and facilitating the identification of biologically meaningful phenotypes.

From a translational perspective, piRNAs may act as mediators linking genetic susceptibility to the actual pathogenic pathways and identifying piRNAs that modulate genes or pathways with high PRS contribution allows the discovery of functional risk biomarkers and potential therapeutic targets.

These findings may enhance our understanding of the complex pathophysiology of CAD and the integrated PRS–piRNA approach has the potential to refine predictive models, enable earlier diagnosis, and support a more precise and biologically informed precision medicine.

However, the clinical translation of piRNA quantification requires careful consideration. Although piRNAs are not yet included in routine screening protocols, several advances support their potential implementation. As costs continue to decrease and analytical pipelines become more streamlined, piRNA profiling may realistically be integrated into centralized clinical laboratories, analogous to current miRNA-based assays. Future work

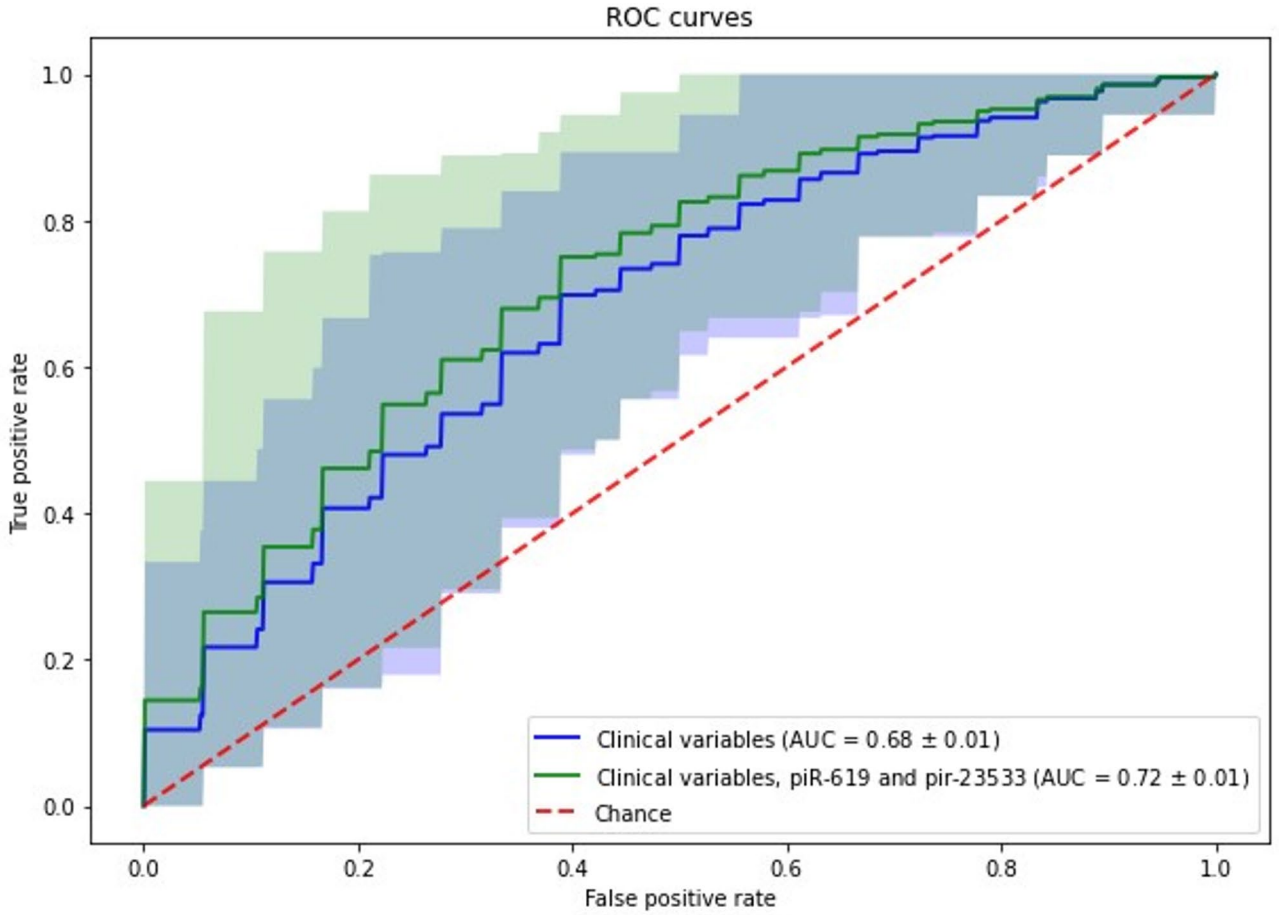

Fig. 5. Receiver Operating Characteristic (ROC) curves for the two predictive models. The first model (blue curve) was fitted using only clinical, anthropometric and lifestyle variables (sex, age, smoking status, alcohol consumption, cholesterol, waist-hip ratio, hypertension, hyperlipidemia treatment). In the second model (green curve) both covariates, piR-619 and piR-23,533 were taken into account. Discriminative performance was evaluated using a Random Forest classification model with repeated stratified 5-fold cross-validation. Area under the ROC curve (AUC), sensitivity, and specificity are reported as mean ± standard error across all folds. Red dashed line is the random classifier. x-axis = False positive rate (1-specificity); y-axis = True positive rate (sensitivity).

should focus on harmonizing pre-analytical procedures, validating piRNA signatures in larger independent cohorts, and evaluating their incremental predictive value in prospective clinical settings.

We are aware that the current study presents some challenges that need to be overcome. While ExoQuick is widely used, it's important to note that it may co-precipitate other molecules, affecting the purity of the isolated EVs.

Additional analyses are needed to further characterize the EVs and identify their cells of origin. As serum is known to be enriched in platelet-derived EVs, and their abundance is further elevated in the context of cardiovascular disease, the potential involvement of these vesicles and their cargo in the pathogenesis of CAD, remain a key area for future research.

Another weakness of this study is the limited number of preclinical CADs, which may not accurately represent the broader population and limit the generalizability of findings. Larger independent cohorts are needed to validate these observations and accurately determine the incremental predictive value of this integrative model, especially considering the profiling of additional molecules layers.

Conclusions

These findings are limited by the sample size and the absence of a replication cohort, and validation in external populations will be essential to confirm the robustness and generalisability of these results. Although we controlled for several clinical factors, other parameters could influence sncRNAs levels, and should be confirmed in independent cohorts which share the main characteristics of the EPICOR study.

Moreover, the integration with PRS paves the way to further investigate the combination of PRS with additional biomarkers as a risk-enhancing factor in individuals at risk of developing CAD.

Further research is warranted to validate our findings on the potential use of piRNAs in identifying individuals at elevated risk for developing CAD alone or in combination with other biomarkers. Such advancements could enable more targeted monitoring and screening protocols, ultimately aiding in the early identification of patients at high CAD risk, as well as to predict disease progression. Our results support the hypothesis that the integration of several biochemical and omics biomarkers can potentially contribute to achieve the goal of a reduction in morbidity and mortality related to these deadly diseases through the adoption of a real omic-based prevention and personalized medicine.

Materials and methods

Study cohort

Study participants included 91 prospective CAD individuals and 91 matched controls (no evidence of cancer or CAD), belonging to the Italian cardiovascular component of The European Prospective Investigation into Cancer and Nutrition (EPIC) study[38].

The EPIC study consists of more than half a million individuals recruited from 10 European countries to investigate the relationships among diet, nutritional status, lifestyle and environmental factors, and the incidence of cancer and other chronic diseases. A workflow that outlines sample selection has been reported in Supplementary Fig. S6.

The Italian cardiovascular component of EPIC (EPICOR study) is a case-cohort study nested in the EPIC-Italy cohort, aimed at investigating the causes of cardiovascular outcomes such as myocardial infarction, acute coronary syndrome, ischemic cardiomyopathy, coronary or carotid revascularization, ischemic- or haemorrhagic stroke[39]. The study sample was selected based on a clinical diagnosis of myocardial infarction, coronary revascularization, or both. Patients with type 2 diabetes were excluded. The cohort comprised 182 paired participants from the EPICOR–Turin study, divided into two clinical groups: 91 preclinical cases who experienced a CAD event an average of 5.95 years after recruitment, and 91 matched healthy controls. Matching was performed according to sex, age, and recruitment center.

All 182 samples were analysed by miRNA sequencing. PRS was calculated for 177 samples (88 CAD cases and 89 controls) due to quality control failures during genotyping.

The study protocol was approved by the Ethics Committee of recruiting centres and all participants gave written informed consent[40,41].

Our study complies with the Declaration of Helsinki principles, and conforms to ethical requirements.

Detailed information on lifestyle was collected through the EPIC lifestyle questionnaire by each participant, including information on previous diseases[38].

We selected etiological risk factors for CAD events from the questionnaire. The assessment of lifestyle (alcohol consumption, smoking habits) and clinical risk factors (hypertension and hyperlipidaemia treatments) were described in Trajkova and colleagues[42]. Briefly, the alcohol consumption was calculated from the questionnaires and categorized according to the quantity of alcohol intake: 0 g/day (non-drinker), 1–12 g/day (light drinkers), and > 12 g/day (heavy drinkers). The waist hip ratio was calculated from the parameters collected at the recruitment.

The smoking status for each participant was defined as never, former smoker and smoker according to the lifetime number of cigarettes and duration of smoking.

Information about blood pressure, body mass index (BMI), hyperlipidemia and hypertension treatments were collected at the recruitment.

Isolation of serum extracellular vesicles and RNA extraction

EVs were isolated from 200 µl of serum using ExoQuick precipitation solution (System Biosciences, USA) according to manufacturer's instructions. Briefly, 200 µl of serum was processed with 50.4 µl ExoQuick solution and stored at 4 °C overnight. The EVs pellets were resuspended with 200 µl of nuclease free water and RNA was immediately isolated from the solution. Total RNA was extracted with miRNeasy serum/plasma kit (Qiagen, Germany) using the QIAcube extractor (Qiagen, Germany) according to manufacturer′s instructions. RNA concentration was determined for all samples with Qubit 2.0 Fluorometer with miRNA assay kit (ThermoFisher, USA).

EV concentration was measured by Nanosight NS300 (Malvern Instruments Ltd., Malvern, UK) as described in Casalone et al. [43]. A representative sample was characterized with Nanoparticle tracking analysis (NTA). The integrity of EVs was assessed by Transmission electron microscopy (TEM) as described in Verta et al. [44].

Briefly, the NTA NS300 (Malvern Instruments Ltd, Malvern, UK) equipped with a 488 nm laser module utilizes Brownian motion and refraction index. The sample was diluted 1:200 in physiologic solution filtered with 100 nm pore size. Three videos of 30s at camera level 15 and threshold 5 were captured. TEM was performed on S-EVs placed on 200-mesh nickel formvar carbon-coated grids (Electron Microscopy Science, Hatfield, PA, USA) and left to adhere for 20 min. The grids were then incubated with 2.5% glutaraldehyde containing 2% sucrose and, after washings in distilled water, the EVs were negatively stained with NanoVan (Nanoprobes, Yaphank, NY, USA) and observed using a Jeol JEM 1010 electron microscope (Jeol, Tokyo, Japan). In Caviglia et al., we also demonstrated with super-resolution microscopy assay the expression of the classical exosomal markers CD63, CD9, and CD81 on the surface of EVs following ExoQuick precipitation[45].

Library preparation and next generation sequencing

Small non-coding RNA libraries were constructed starting from 6 ul of total RNA, using NEBNext Multiplex Small RNA Library Prep set for Illumina (New England Biolabs, Inc., USA). The cDNA libraries were purified with the Qiagen PCR Purification kit following the modifications indicated by the manufacturer′s protocol. Six pools of cDNA barcoded samples were prepared and each pool was finally size-selected sncRNAs (miRNAs

and piRNAs) in a 6% PolyAcrylamide Gel (ThermoFisher, USA). Fragments with an insert of 140 nucleotides, corresponding to miRNAs, and of 150 nucleotides, corresponding to piRNAs, were cut out and purified with Qiagen Gel Extraction MiniElute columns (Qiagen, Germany) following the modifications indicated in NEBNext Multiplex Small RNA Library Prep Protocol. A BioAnalyzer 2100 system (Agilent Technologies, USA) was used to check the size of each cDNA library fragment. Single-end sequencing (75 nt) was performed on the NextSeq550 platform (Illumina, USA).

Sequencing data analysis

The data from small RNA sequencing were computationally processed and analyzed. Briefly, adapter removal was performed using Cutadapt[46] with the following options:

cutadapt --adapter=AGATCGGAAGAGCACACGTCT --minimum-length = 14 --overlap = 10 --trim-n. --error-rate = 0.1.

Reads shorter than 14 nucleotides after trimming were discarded. The rest were mapped in a single pass to a joint reference including miRBase v.22[47] and piRBase v1.0[48], using the BWA aligner (v0.7.17)[49] with the following options:

bwa mem -t 8 -T 14 -A 1 -k 10 -a.

In particular we reduced the seed length to 10 bases and the minimum output score to 14, to guarantee that shorter queries were correctly aligned and reported. All possible alignments were included in the output bam files.

Alignments were then filtered with a custom Python script that discarded unsatisfactory alignments and selected the best one for each read in case of multiple alignments passing the filter. In detail, unaligned or reverse mapped reads were discarded. Then, after parsing the CIGAR and some optional tags reported by BWA, we discarded those alignments with indels or mismatches and those where the matching bases were less than 85% of the query or less than 70% of the reference sequence. We also dropped those that had more than 2 bases of left clipping. This was done to check that the alignment is close to a global match, e.g., avoiding partial matches from miRNA into a longer piRNA. We also reasoned that for very short sequences like these, mismatches or indels due to errors are unlikely and could instead produce errors. After the filtering, if there are still multiple alignments for the same read, we pick the best one to assign the read to a unique miRNA or piRNA, prioritising those where the matching sequence comprises the whole query and/or reference sequences.

We reported in Supplementary Table S7 the raw count matrix mapped across all samples.

Technical validation by RT-qPCR

We validated the differentially expressed sncRNAs by quantitative reverse transcription polymerase chain reaction (RT-qPCR) in 20 randomly selected pre-clinical cases and their 20 healthy matched controls. To identify the candidate sncRNAs, FDR < 0.05 and |log2FoldChange(FC)|>0.6 were used as selection criteria. For technical confirmation of the sncRNA-seq results, RT-qPCR was performed using miRCURY LNA miRNA SYBR Green PCR using miRCURY LNA miRNA Custom PCR Panels (QIAGEN Sciences, Germany), with CFX96 Real-time PCR Detection System (Bio-Rad, US). The sequences corresponding to each validation target (hsa-miR-451a, hsa-miR-375-3p, hsa-miR-194-5p, hsa-miR-22-5p, hsa-piR-20266, hsa-piR-23533, hsa-piR-27282, hsa-piR-28212, hsa-piR-1043, and hsa-piR-619), together with their total read counts across all samples, are listed in Supplementary Table S8. For validation, the canonical or predominant form of each sequence was selected. The full sncRNA sequence of miRNAs and piRNAs assay has been reported in Supplementary Table S9.

Real-time PCR was performed on CFX96 Real-Time PCR machine (BioRad, USA) according to the manufacturer's protocol. Briefly, 2 µl of total RNA was used for Universal RT using mirCURY LNA RT Kit (QIAGEN Sciences, Germany). The cut-off was set at 38 Cq (the PCR cycle number at which the sample′s reaction curve intersects the threshold line), No Template Controls (NTCs) did not amplify or the difference between NTCs Cq value and sample Cq value was at least 10. Unisp6 RNA spike-in was used as reverse transcription positive control in PCR panels and added while preparing the RT master mix as the protocol suggested. Moreover, Unisp3 was used as an interplate calibrator.

The cDNA was diluted with nuclease-free water (1:40), 4 µl of product was then used for Real-time PCR using miRCURY SYBR Green PCR Master Mix (QIAGEN Sciences, Germany). Real-time PCR was performed on CFX96 Real-Time PCR machine (BioRad, USA) according to the manufacturer′s protocol.

The custom plates were pre-spotted with miRNA primers, and all assays were performed in single. Both the reaction mix and DNA were dispensed using an automated liquid-handling system (Myra, Resnova).

Data acquisition and subsequent data analyses were performed using the CFX Manager software (Bio-Rad, Hercules, CA, USA). Relative fold changes in sncRNA expression were calculated from normalized Cq values using miR-103a-3p as reference, selected for its minimal variability and stable expression across groups are reported in Supplementary Table S10.

Functional enrichment analysis of MiRNAs target genes

Validated target genes were retrieved using the MultiMiR Bioconductor package (github.com/KechrisLab/multiMiR)[50], which integrates data from miRecords, miRTarBase, and TarBase. The analysis was conducted in RStudio (2022.07.0). Only those genes with a binding probability of at least 95% were considered as targets and taken into account for further analyses.

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Gene Ontology (GO) Biological process, implemented respectively in ShinyGO (v0.741) (bioinformatics.sdstate.edu) and in WEB-based GEne SeT AnaLysis Toolkit (WebGestalt, webgestalt.org), were used as reference for the pathway enrichment analyses and potential functional roles associated with the genes targeted by validated miRNAs[19].

Functional enrichment analysis for piRNAs was not performed due to the lack of available human target genes databases.

CAD polygenic risk score estimation

Genetic data was collected through DNA analysis by whole-genome genotyping microarrays provided by Illumina. Data imputation was made through the Michigan Imputation server, using the Genome Reference Consortium Human Build 37, r1.1 2016 as reference panel. We filtered data after imputation considering as criteria a minor allele frequency (MAF) < 0.01 and an R2 > 0.6, to obtain a total of 7,080,677 variants.

PGS000329 was published by Mars N. et al. and we downloaded the scoring file from PGS Catalog (https://www.pgscatalog.org/score/PGS000329/) [51]. This score was developed from a GWAS of 408,458 European individuals as our cohort. Then we checked that more than 95% of variants contained in PGS000329 had a good imputation score (R2 > 0.8), to verify how reliable the polygenic score could be.

The PRS for each individual were calculated through the Michigan Imputation Server as for the Ancestry Estimation and Principal Component Analyses to verify their geographical origin.

We used the Kolmogorov-Smirnov test to compare the scores distribution and the Mann–Whitney test to determine differences between groups. Pearson correlation was used to measure linear correlation between variables. Odds Ratios (OR) with 95% CI were calculated keeping as reference all individuals with a score lower than 80th percentile and using questionr (version 0.7.8) R package[52].

Statistical analysis

To analyze the sncRNAs expression data obtained from NGS, DESeq2 Bioconductor package (v.1.22.2)[53] was used. The differential expression analysis was adjusted for sex, age, smoking status, alcohol consumption, cholesterol, waist-hip ratio, hypertension (including both individuals under treatment and untreated hypertensive subjects), hyperlipidemia treatment, and potential batch effects arising from differences between sequencing runs, which were balanced according to the matching criteria. All cases were free of clinically diagnosed CAD at the time of serum collection and developed the event on average 5.9 years later.

Samples with missing covariates were dropped and only sncRNAs with a median expression of at least ten reads were included. SncRNAs obtained from the NGS analysis were considered significantly differentially expressed between pre-diagnostic cases and matched controls if their p-value was below the 0.05 threshold after adjustment for multiple testing by false discovery rate (FDR).

Boxplots representing the normalized relative expression (read counts reported in log2 scale) of selected sncRNAs in pre-clinical CAD cases compared to healthy controls were retrieved with RStudio (2022.07.0) using the ggplot2 package[54].

Correlation analyses between log-transformed, normalised NGS read counts and RT-qPCR ΔCt values for each sncRNA across samples were performed in R (v4.1.3). Depending on the distributional properties of each dataset, either Pearson's parametric correlation or Spearman's non-parametric rank correlation was applied, adopting a significance threshold of $p \leq 0.05$.

Differential expression analysis for each of the tested sncRNAs in RT-qPCR was performed in a generalized linear model (GLM) analysis adjusted for the same clinical variables described previously.

In the absence of a replication cohort, we assessed the discriminative power of validated piRNAs in distinguishing pre-clinical CAD cases from healthy controls using a Random Forest model applied to the same dataset from the EPICOR study.

We have evaluated the model with 1000 trees and a maximum depth of four, using repeated stratified 5-fold cross-validation, 10 times with reshuffling and reporting the mean result across all folds from all runs. Scores are reported as averages followed by their standard deviation.

Two models were fitted, one using only clinical variables (sex, age, smoking status, alcohol consumption, cholesterol, waist-hip ratio, hypertension, hyperlipidaemia treatment) and one with covariates and sncRNAs. The average and the standard error were calculated for the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve and specificity and sensitivity were calculated with a standard cut-off value of 0.5.

Data availability

The genetic and sequencing samples data that support the findings of this study are available from the European Prospective Investigation into Cancer and Nutrition (EPIC) project, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. The produced data are available from Prof. Giuseppe Matullo (giuseppe.matullo@unito.it) upon reasonable request and with permission of the EPIC committee. Raw count matrix of the aligned small RNAs across all sequenced samples is provided in Supplementary material.

Received: 21 January 2025; Accepted: 29 January 2026
Published online: 11 February 2026

References

1. Roth, G. A. et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J Am. Coll. Cardiol.* **22** (25), 2982–3021 (2020).
2. Hajar, R. Risk factors for coronary artery disease: historical perspectives. *Heart Views.* **18** (3), 109–114 (2017).
3. Marston, N. A. et al. Predictive utility of a coronary artery disease polygenic risk score in primary prevention. *JAMA Cardiol.*. **1** (2), 130–137 (2023).
4. Loche, E. & Ozanne, S. E. Early nutrition, epigenetics, and cardiovascular disease. *Curr Opin. Lipidol.* **27** (5), 449–458 (2016).

5. Zhang, X. & Schulze, P. C. MicroRNAs in heart failure: Non-coding regulators of metabolic function. *Biochim Biophys. Acta* **1862** (12), 2276–2287 (2016).
6. Romaine, S. P., Tomaszewski, M., Condorelli, G. & Samani, N. J. MicroRNAs in cardiovascular disease: an introduction for clinicians. *Heart*. **101** (12), 921–928 (2015).
7. Mujwara, D. et al. Integrating a polygenic risk score for coronary artery disease as a risk-enhancing factor in the pooled cohort equation: a cost-effectiveness analysis study. *J Am. Heart Assoc.*. **21** (12), e025236 (2022).
8. Ai, J. et al. Circulating microRNA-1 as a potential novel biomarker for acute myocardial infarction. *Biochem. Biophys. Res. Commun.*. **1** (1), 73–77 (2010).
9. Devaux, Y. et al. Diagnostic and prognostic value of Circulating MicroRNAs in patients with acute chest pain. *J. Intern. Med.*. **277** (2), 260–271 (2015).
10. Peters, L. J. F. et al. Small things matter: relevance of MicroRNAs in cardiovascular disease. *Front. Physiol.* **11**, 793 (2020).
11. Zeng, Q. et al. PIWI-interacting RNAs and PIWI proteins in diabetes and cardiovascular disease: molecular pathogenesis and role as biomarkers. *Clin Chim. Acta*. **518**, 33–37 (2021).
12. Rayford, K. J. et al. PiRNAs as modulators of disease pathogenesis. *Int. J. Mol. Sci.* **27** 22 (5). (2021).
13. Rajan, K. S. et al. Abundant and altered expression of PIWI-Interacting RNAs during cardiac hypertrophy. *Heart Lung Circ.* **25** (10), 1013–1020 (2016).
14. Zhong, N., Nong, X., Diao, J. & Yang, G. piRNA-6426 increases DNMT3B-mediated SOAT1 methylation and improves heart failure. *Aging (Albany NY)* **30** (6), 2678–2694 (2022).
15. Chen, B. et al. Targeting a cardiac abundant and fibroblasts-specific PiRNA (CFRPi) to attenuate and reverse cardiac fibrosis in pressure-overloaded heart failure. *Transl Res.*. **267**, 10–24 (2024).
16. Borgovan, T., Crawford, L., Nwizu, C. & Quesenberry, P. Stem cells and extracellular vesicles: biological regulators of physiology and disease. *Am J. Physiol. Cell. Physiol.*. **1** (2), C155–C166 (2019).
17. Xiao, F. et al. MiRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*. **37** (Database issue), D105–110 (2009).
18. Karagkouni, D. et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.*. **4** (D1), D239–D245 (2018).
19. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. **15** (8), 2628–2629 (2020).
20. Sessa, F., Salerno, M., Esposito, M., Cocimano, G. & Pomara, C. MiRNA dysregulation in cardiovascular diseases: current opinion and future perspectives. *Int. J. Mol. Sci.* **8** ;24(6). (2023).
21. Satam, H. et al. Next-Generation sequencing technology: current trends and advancements. *Biology (Basel)* **13** 12(7). (2023).
22. Zhu, M. et al. Ischemic postconditioning protects remodeled myocardium via the PI3K-PKB/Akt reperfusion injury salvage kinase pathway. *Cardiovasc. Res.* **1** (1), 152–162 (2006).
23. Chen, L., Bai, J., Liu, J., Lu, H. & Zheng, K. A Four-MicroRNA panel in peripheral blood identified as an early biomarker to diagnose acute myocardial infarction. *Front. Physiol.* **12**, 669590 (2021).
24. Hu, H. et al. PiR-hsa-23533 promotes malignancy in head and neck squamous cell carcinoma via USP7. *Transl Oncol. Jul.* **45**, 101990 (2024).
25. Yang, J. et al. Cardiac fibroblasts-specific USP7 drives post-infarction cardiac fibrosis by deubiquitinating Kruppel-like factor 7 to promote myofibroblast activation. *J. Mol. Cell. Cardiol.*. **17**, 210 109–126 (2025).
26. Blazejowska, E. et al. Diagnostic and prognostic value of MiRNAs after coronary artery bypass grafting: A review. *Biology (Basel)* **19** 10(12). (2021).
27. Ghafouri-Fard, S., Gholipour, M. & Taheri, M. Role of MicroRNAs in the pathogenesis of coronary artery disease. *Front. Cardiovasc. Med.* **8**, 632392 (2021).
28. Taraldsen, M. D., Wiseth, R., Videm, V., Bye, A. & Madssen, E. Associations between Circulating MicroRNAs and coronary plaque characteristics: potential impact from physical exercise. *Physiol. Genomics*. **1** (4), 129–140 (2022).
29. Marques, F. Z., Vizi, D., Khammy, O., Mariani, J. A. & Kaye, D. M. The transcardiac gradient of cardio-microRNAs in the failing heart. *Eur. J. Heart Fail.*. **18** (8), 1000–1008 (2016).
30. Yang, J., Xue, F. T., Li, Y. Y., Liu, W. & Zhang, S. Exosomal PiRNA sequencing reveals differences between heart failure and healthy patients. *Eur Rev. Med. Pharmacol. Sci.* **22** (22), 7952–7961 (2018).
31. Zhu, X. et al. miR-194 inhibits the proliferation, invasion, migration, and enhances the chemosensitivity of non-small cell lung cancer cells by targeting forkhead box A1 protein. *Oncotarget* **15** (11), 13139–13152 (2016).
32. Pang, Q. et al. Regulation of the JAK/STAT signaling pathway: the promising targets for cardiovascular disease. *Biochem Pharmacol.*. **213**, 115587 (2023).
33. Mascareno, E. et al. JAK/STAT signaling is associated with cardiac dysfunction during ischemia and reperfusion. *Circulation*. **17** (3), 325–329 (2001).
34. Gay, A. & Towler, D. A. Wnt signaling in cardiovascular disease: opportunities and challenges. *Curr Opin. Lipidol.* **28** (5), 387–396 (2017).
35. Jian, W. et al. Association between serum HER2/ErbB2 levels and coronary artery disease: a case-control study. *J Transl Med.* **11** (1), 124 (2020).
36. Sciarretta, S., Forte, M., Frati, G. & Sadoshima, J. New insights into the role of mTOR signaling in the cardiovascular system. *Circ Res.*. **2** (3), 489–505 (2018).
37. Debernardi, C. et al. Population heterogeneity and selection of coronary artery disease polygenic scores. *J Pers. Med.* **26** ;14(10). (2024).
38. Palli, D. et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy prospective Study. Design and baseline characteristics of participants. *Tumori* . **89** (6), 586–593 (2003).
39. Bendinelli, B. et al. Fruit, vegetables, and Olive oil and risk of coronary heart disease in Italian women: the EPICOR study. *Am J. Clin. Nutr.*. **93** (2), 275–283 (2011).
40. Di Castelnuovo, A. et al. Elevated levels of D-dimers increase the risk of ischaemic and haemorrhagic stroke. Findings from the EPICOR study. *Thromb Haemost.*. **112** (5), 941–946 (2014).
41. Sieri, S. et al. Dietary glycemic load and index and risk of coronary heart disease in a large Italian cohort: the EPICOR study. *Arch Intern. Med.*. **12** (7), 640–647 (2010).
42. Trajkova, S. et al. Impact of preventable risk factors on stroke in the EPICOR study: does gender matter? *Int J. Public. Health* **62** (7), 775–786 (2017).
43. Casalone, E. et al. Serum extracellular Vesicle-Derived MicroRNAs as potential biomarkers for pleural mesothelioma in a European prospective study. *Cancers (Basel)* **25** 15 (1). (2022).
44. Verta, R. et al. Generation of Spike-Extracellular vesicles (S-EVs) as a tool to mimic SARS-CoV-2 interaction with host cells. *Cells* **3** 11 (1). (2022).
45. Caviglia, G. P. et al. Extracellular vesicles mirnome profiling reveals MiRNAs engagement in dysfunctional lipid Metabolism, chronic inflammation and liver damage in subjects with metabolic dysfunction-associated steatotic liver disease. *Aliment Pharmacol. Ther. Jul.* **62** (1), 22–32 (2025).
46. M., M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17** (1), 10–12 (2011).
47. Griffiths-Jones, S. MiRBase: MicroRNA sequences and annotation. *Curr Protoc. Bioinf.* **12** 19 11–12 19 10.

48. Wang, J. et al. PiRBase: integrating PiRNA annotation in all aspects. *Nucleic Acids Res.* **7** (D1), D265–D272 (2022).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **15** (14), 1754–1760 (2009).
50. Ru, Y. et al. The MultiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res.* **42** (17), e133 (2014).
51. Mars, N. et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med.*. **26** (4), 549–557 (2020).
52. Larmarange JBaFBaJ. questionr: Functions to make surveys processing easier.
53. Love, M. I., Huber, W. & Anders, S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15** (12), 550 (2014).
54. Wickham, H. ggplot2: Elegant graphics for data analysis.

Acknowledgements
The authors acknowledge Giovanni Camussi (Dep of Medical Sciences, University of Turin) for his contribution in characterizing extracellular vesicles.

Author contributions
E.C. G.M. A.A. C.S. conceptualised the study. E.C. G.M. A.A. and C.C initiated and designed the experiments. G.M. P.V. and C.S. help in the selection of the samples. E.C. M.R. and C.D. wrote the main manuscript text. G.B. S.A. C.D. C.D.P. E.C. performed the bioinformatic and statistical analyses. A.P. C.G. and G.A. revised the manuscript.All authors gave final approval of the version to be published. All authors made a significant contribution to the work reported.

Funding
EPIC, EPICOR and EPICOR2 projects were supported by the Compagnia di San Paolo for the EPIC, EPICOR and EPICOR2 projects (SP, SS, RT, PV, LI, CS, GM) and from the Ministero dell'Istruzione, dell'Università e della Ricerca—c (n° D15D18000410001, to G.M.) to the Department of Medical Sciences, University of Torino. This work was also supported by the "Genoma mEdiciNa pERsonalizzatA – GENERA", funded from the Ministero dell'Istruzione, dell'Università e della Ricerca 2021 (n° D73C22000960001) and by CARdiomyopathy in type 2 DIAbetes mellitus (Cardiateam) funded from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 821508.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
The European Prospective Investigation into Cancer and Nutrition (EPIC) study protocol was approved by the ethics committees centralized at the International Agency for Research on Cancer (Lyon, France) (reference number: IEC 24 – 08). The EPICOR Study, a case–cohort study nested within the EPIC–Italy prospective cohort, was approved by the HuGeF Ethics Committee in Turin on 15 December 2010.

Informed consent
Informed consents for the genetic research were obtained from all subjects involved in both the studies cited earlier.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-38355-4.

Correspondence and requests for materials should be addressed to E.C. or G.M.