Biomedicine & Pharmacotherapy 139 (2021) 111621

Contents lists available at ScienceDirect

Biomedicine & Pharmacotherapy

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

Purine metabolite-based machine learning models for risk prediction, prognosis, and diagnosis of coronary artery disease

Sunhee Jung [a,1], Eunyong Ahn [a,1], Sang Baek Koh [b], Sang-Hak Lee [c], Geum-Sook Hwang [a,d,*]

[a] Integrated Metabolomics Research Group, Western Seoul Center, Korea Basic Science Institute, 150 Bugahyeon-ro, Seodaemun-gu, Seoul 03759, South Korea
[b] Department of Preventive Medicine, Yonsei University Wonju College of Medicine, Wonju 26426, South Korea
[c] Division of Cardiology, Department of Internal Medicine, Severance Hospital, Yonsei University College of Medicine, 134 Shinchon-dong, Seodaemun-gu, Seoul 03722, South Korea
[d] Department of Chemistry and Nano Science, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, South Korea

ARTICLE INFO

Keywords:
Coronary artery disease (CAD)
Metabolomics
Machine learning model development
Purine metabolism
Biomarker

ABSTRACT

Alterations in xanthine oxidase activity are known to be pathologically influential on coronary artery disease (CAD), but the association between purine-related blood metabolites and CAD has only been partially elucidated. We performed global metabolomics profiling and network analysis on blood samples from the Wonju and Pyeongchang (WP) cohort study ($n = 2055$) to elucidate the importance of purine related metabolites associated with potential CAD risk. Then, 5 selected serum metabolites were quantified from the WP cohort, Shinchon cohort ($n = 259$), and Shinchon case control ($n = 424$) groups to develop machine learning models for 10-year risk prediction, relapse within 10 years and diagnosis of the disease via 100 repeated 5-fold cross-validations of logistic models. The combination of purine metabolite levels or only xanthine levels in blood could be applied for machine learning model development for major adverse cardiac and cerebrovascular event (MACCE, cerebrovascular death, nonfatal myocardial infarction, percutaneous transluminal coronary angioplasty, coronary artery bypass graft, and stroke) risk prediction, relapse of MACCEs among patients with myocardial infarction history and diagnosis of stable CAD. In particular, our research provided initial evidence that blood xanthine and uric acid levels play different roles in the development of machine learning models for primary/secondary prevention or diagnosis of CAD. In this research, we determined that purine-related metabolites in blood are applicable to machine learning model development for CAD risk prediction and diagnosis. Also, our work advances current CAD biomarker discovery strategies mainly relying on clinical features; emphasizes the differential biomarkers in first/secondary prevention or diagnosis studies.

1. Introduction

Cardiovascular disease (CVD) is one of the leading causes of death worldwide, also in Korea [1,2]. CVDs are multifactor diseases that are influenced by a wide range of genetic, dietary, and environmental factors [3]. As CVDs are complex pathophysiological diseases, further, intensive research is needed to elucidate the risk prediction and diagnosis for improvements in outpatient and inpatient care [4,5]. Currently, various methodologies, such as electrocardiographic findings, imaging tools, and elevated biomarkers of myocardial necrosis, are used for CVD diagnosis [6]. Metabolomics is considered an emerging technology to identify metabolic biomarkers for CVD and coronary artery disease (CAD), which is the most common type of CVD including myocardial infarction (MI) and angina [7]. It has been reported that the metabolism of amino acids, acyl-carnitines, and purines is significantly altered in patients with MI compared to healthy individuals [6]. The metabolite levels in glycolysis and fatty acid metabolism have been found to be changed in the plasma from atherosclerosis and MI patients [8,9]. Although previous studies have reported that various metabolites are linked to the diagnosis or potential risk of CAD, prospective metabolomics studies in which a large number of normal participants are enrolled at baseline are very limited in terms of CVD.

There have been several attempts to utilize metabolomics for the diagnosis of CADs. For instance, a machine learning classification

* Corresponding author at: Integrated Metabolomics Research Group, Western Seoul Center, Korea Basic Science Institute, 150 Bugahyeon-ro, Seodaemun-gu, Seoul 03759, South Korea.
E-mail address: gshwang@kbsi.re.kr (G.-S. Hwang).
[1] These authors are equally contributed.

https://doi.org/10.1016/j.biopha.2021.111621
Received 29 January 2021; Received in revised form 23 March 2021; Accepted 12 April 2021
Available online 10 May 2021

algorithm using metabolites as features predicted spontaneous MI patients from patients undergoing diagnostic coronary angiography [10]. Additionally, a recent CAD study was characterized by distinct metabolic perturbations that were used as biomarkers to distinguish various types of CADs, including atherosclerosis, angina, and MI [11].

A case-cohort study partially elucidated the metabolomic profiling strategy's role in CAD risk prediction [3,12]. High levels of metabolite species, including amino acids, acylcarnitines, triglycerides, cholesterol esters, and phosphatidylethanolamines, were associated with potential CVD risk [3,12]. Canela et al. evaluated the predictive power of branched-chain amino acids (BCAAs), such as leucine, isoleucine, and valine, in a case-cohort study. After 1 year of follow-up of 970 participants, hazard ratios (HRs) for the 4th quartile of baseline BCAA concentrations were higher than those for the 1st quartile, indicating association of BCAA concentrations with CVD events including stroke, MI, or death [12].

In fact, age and diabetes are also known as major risk factors for developing CVDs, and clinical risk factor-associated metabolic features have been introduced [13,14]. In retrospective analysis, plasma amino acids were associated with the onset of severe CVDs in diabetic patients; 63 patients developed CVDs during an observational 10-year follow-up study of 420 diabetic patients [15]. The Bruneck study in Italy is a well-planned, large population or community-based survey study that enables reliable evaluation of biomarker discovery for metabolic disease [16]. Using these data, Stegemann et al. observed diverse cardiovascular effects on lipid species in plasma [17]. There were strong associations between CVD events and the levels of CEs and TAGs, with low carbon numbers or a low double-bond content [17,18]. Although these studies suggested lipid features' possible role in the prevention of CADs, they could not suggest biomarkers from aqueous metabolic features.

Moreover, there have been no systematic attempts to compare different types of metabolomic biomarker discoveries for risk prediction, prognosis, and diagnosis of disease by controlling analytical platforms, data processing, and machine learning algorithms. Moreover, notably, there have been few attempts to quantify aqueous metabolites using large-sized cohort data with a normal baseline.

Here, we conducted liquid chromatography–mass spectrometry (LC-MS)-based metabolic profile analysis to specify the aqueous metabolites applicable to advance CAD patient care followed by absolute quantification of the targeted metabolites in large-sized population-based survey data with normal baseline. Our work enabled us to discover aqueous serum metabolite biomarkers applicable for CAD risk prediction before the onset of disease. A targeted metabolomics approach was conducted in the analysis of three different datasets, which aimed to develop machine learning models to predict the potential risk, prognosis, and diagnosis of CAD. Notably, our research first introduces the possibility and potential limitations of applying metabolic biomarker analysis for risk prediction to develop biomarkers for prognosis or diagnosis using three equivalently processed datasets.

2. Material and methods

2.1. Study design

The WP cohort study was led by the Wonju Severance Christian Hospital in Korea. This study was designed as a large cohort to develop a CAD risk prediction model in the Korean population. In this study, 2055 people (247 with MACCEs, 1808 without MACCEs) enrolled between December 2005 and January 2007 for cohort tracking until 2015. Blood plasma from participants was collected at the beginning of cohort tracking, and participants' clinical parameters were simultaneously assessed.

The SC cohort study, led by Yonsei Medical University, included 259 CAD patients with either stable CAD or MI. The patients were recruited from January 2006 until February 2014, and their medical records were tracked for up to 11 years. During enrollment in the study, serum samples were obtained for metabolomics analysis, and participants' clinical parameters were assessed. Further CAD events were defined according to MACCE.

A total of 150 patients with stable CAD and 274 counterpart controls were recruited for the case control study led by Yonsei Medical University. The participants' samples were collected from November 2000 to March 2011. Similar to other studies, blood serum samples from participants were collected for metabolomics analysis, and participants' clinical parameters were assessed.

2.2. Sample randomization

Due to a large number of samples, we split the sample sets into 3 analysis batches according to the sample collection sites. Samples in each data set were randomized for subsequent analysis.

2.3. Sample extraction

To extract metabolites in serum, 50 μL of serum sample was mixed with 500 μL of chloroform/methanol (2:1, v/v) and 100 μL of water. The mixture was vortexed and incubated at 4 °C for 10 min. After centrifugation, aliquots of the supernatants were dried in a vacuum concentrator. The extracts were diluted with 100 μL of acetonitrile/water mixture (75:25, v/v) for global profiling and 100 μL of acetonitrile/water mixture (20:80, v/v) for targeted profiling.

2.4. Global profiling for primary prevention of CAD disease

For the primary prevention study, LC- electrospray ionization(ESI)-MS/MS analyses of sample extracts were performed on a triple TOF™ 5600 MS/MS System (AB Sciex, Canada) combined with a UPLC system (Waters, USA). LC separations were carried out on a ZIC®-HILIC column (2.1 mm × 100 mm, 3.5 μm; SeQuant). The column temperature and flow rate were set to 35 °C and 0.4 mL/min, respectively. The mobile phases used were 10 mM ammonium acetate and 0.1% formic acid water:acetonitrile (5:95 v/v) (A) and in water:acetonitrile (50:50 v/v) (B). The linear gradient program was as follows: 1% B from 0 to 2 min, 1–55% B from 2 to 8 min, 55–99% B from 8 to 9 min, 99% B from 9 to 11 min, 99–1% B from, 11–11.1 min, and 99% B from 11.1 to 15 min. The injection volume of the sample was 5 μL using partial loop mode for both positive and negative ionization polarity modes. Quality control (QC) samples, which were pooled identical aliquots of the samples, were measured regularly throughout the run for data reproducibility. The spectral data were analyzed by MarkerViewTM (AB Sciex, Canada), which was used to find peaks, perform peak alignment, and generate peak tables of m/z and retention times (min). Spectra were normalized using the total area of the spectra. To identify reliable peaks and remove instrumental bias, peaks with an intensity in QC samples lower than those in a blank sample were eliminated, and peaks with coefficients of variation below 20 in QC samples were selected. Metabolites were tentatively identified by comparing the experimental data against various databases, including the METLIN (metlin.scripps.edu), Human Metabolome (www.hmdb.ca), and MassBank (www.massbank.jp) databases. Fragment patterns (MS/MS spectra) were also used to identify metabolites.

2.5. Quantitative analysis of metabolites in purine metabolism

To quantify the level of metabolites, liquid chromatography-mass spectrometry was performed on an Agilent 1290 Infinity LC and an Agilent 6490 Triple Quadrupole MS system equipped with an Agilent Jet Stream ESI source (Agilent Technologies, USA). MassHunter Workstation (Ver B.06.00, Agilent Technologies, USA) software was used for data acquisition and analysis. LC separations were carried out on a reversed-phase Scherzo SM-C18 column (100 mm × 2 mm, particle size 3 μm, Imtakt, USA) over 12 min. The column temperature and flow rate

were set to 35 °C and 0.4 mL/min, respectively. The binary gradient system comprised 0.1% formic acid in water (solvent A) and methanol (solvent B). The linear gradient used for elution and equilibrating the initial gradient for subsequent runs was as follows: 1% B from 0 to 1.5 min, 7–90% B from 1.5 to 8 min, 90% B from 9 min, 90–7% B from 9 to 10 min, and 7% B from 12. The autosampler temperature was maintained at 4 °C, and the injection volume of the sample was 1 μL. The column effluent was introduced into a TQ mass detector operating in positive electrospray ionization mode. Samples were analyzed in single reaction monitoring (SRM) mode. The SRM transitions were performed with the following operational parameters: gas temperature, 120 °C; gas flow, 11 L/min; nebulizer gas pressure, 40 psi; sheath gas temperature, 350 °C; sheath gas flow, 12 L/min; capillary voltage, + 3.5 kV and − 4.0 kV. The fragmentor voltage was 380 V for all compounds. Samples were split into batches that included 50 samples per day. In addition, to assess the correction of systematic between-batch effects, the same lot number of formic acid and columns were used. All samples were analyzed with identical time sequences each day. The concentration of metabolites was quantified based on a series of external calibration curves generated daily for each batch with authentic standards. The results for the calibration curve were confirmed to be consistent every day, and then the samples were analyzed. Samples that were out of the calibration curve were diluted and reanalyzed. A total of 69 batches were analyzed, including reanalysis due to instrument stability. Data from further analysis were included in a total of 59 batches.

2.6. Pathway analysis

Pathway analysis for global profiling results of the WP cohort was conducted through MetaboAnalyst [19]. Metabolites with *p*-values less than 0.05 from *t*-tests were used as input for the analysis. A hypergeometric test was conducted to evaluate the pathway enrichments based on *Homo sapiens* data in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [20].

2.7. Network analysis

Unmatched control data were separately processed to build correlation coefficient-based network analysis. Quantified purine-related metabolites and assessed clinical parameters are represented as nodes in the network graph. The resampling strategy-based probabilistic context likelihood of relatedness (PCLR) algorithm [21] and R software [22] were applied to determine edges between nodes. The built network was further processed by using Cytoscape [23]. The width and height of nodes were scaled based on stress centrality measures. Stress centrality is the number of shortest paths passing through as a proxy measure to capture the importance of features [24].

2.8. Statistical analysis

We conducted descriptive analysis for each data set. The Wilcoxon rank sum test or T-test was used to evaluate significance for 5 clinical variables and 5 serum metabolite levels. In feature selection of the machine learning procedure, to equivalently process three different studies, 5 common parameters (age, sex, hypertension, hyperlipidemia, and diabetes) were used for model development.

Multivariable-adjusted ORs and 95% CIs were estimated using logistic regression and reported per the standard deviation. In the feature selection procedure, the cross-validation error (CVerror) from 100 repeated 5-fold cross-validations was exhaustively calculated to find the best logistic model. Based on the mean AUC from logistic regression by using 5-fold cross validations, the performance of each model was re-evaluated.

A Cox proportional hazard model was used for survival analysis, and the log rank test was performed to compare the survival plots. The resampling strategy-based PCLR algorithm [21] and R software [22] were applied to determine edges between nodes. R 1.2.1335 software and relevant packages were utilized for all the statistical analyses in this research [22].

2.9. Ethics committee approval and patient consent

All participants provided written consent before the outset of the study. The protocol was approved by the Institutional Review Board (CR105024–026) of Wonju Severance Christian Hospital and the Institutional Review Board (4-2001-0039) of Yonsei Medical University.

3. Results

3.1. Study design and patient information

The data from three different CAD disease studies, the Wonju and Pyeongchang (WP) population-based cohort study, Shinchon (SC) case cohort study, and Shinchon (SC) case control study, were equivalently collected, quantified and analyzed to understand the possible roles of metabolic features in improving CAD patient care. In this study, three datasets were applied to explore the primary and secondary prevention and diagnosis of CAD (Fig. 1). During enrollment, patient blood samples were collected for metabolomics analysis, and clinical parameters were simultaneously assessed (Tables 1–3). A total of 2738 blood samples from all three datasets were analyzed in 59 batches for three months of LC-MS operation.

3.2. Purine metabolites as CAD risk prediction biomarkers

WP population data, the largest data set among the three study groups, were used for discovery of altered metabolic pathways attributed to CAD. While 2055 participants were tracked for approximately 10 years, 247 people had major adverse cardiac and cerebrovascular events (MACCEs). Using the plasma samples from the participants, nontargeted metabolic profiling was performed to identify differential metabolic features followed by metabolic annotation. After peak alignment and removal of missing values of ultra-high performance liquid chromatography/quadrupole time-of-flight mass spectrometry (UPLC/QTOF MS) data, 5109 and 4421 features were detected in positive and negative mode, respectively. A total of 62 metabolites were identified by comparing open source databases.

Based on our following pathway analysis, purine metabolism was identified as the most differentially altered pathway by applying the analysis module in Metaboanalyst 4.0 (Supplemental Fig. 1) [19]. The relative abundances of 4 related metabolites, hypoxanthine, inosine, xanthine, and xanthosine, were associated with MACCEs during the 10-year follow-up period compared to those from normal counterparts. Next, 4 identified purine-related metabolites from our global profiling analysis and uric acid were further measured via quantification analysis based on triple quadrupole (TQ)-MS (Table 4). In this large cohort study, we found that uric acid and xanthosine were increased and hypoxanthine and inosine were decreased in the MACCE group compared to the non-MACCE group without adjusting for other clinical factors. Considering that the two groups significantly differed in several baseline characteristics, our descriptive analysis of purine-related metabolite levels in blood should be further re-evaluated by our following machine learning analysis.

3.3. Building CAD risk prediction model

Using quantified purine metabolite concentrations and assessed clinical features, we performed machine learning analysis to build a linear model for CAD risk prediction. Since the number of samples with events was much smaller than the number of normal samples and the baseline clinical characteristics were not balanced, the same number of randomly matched normal samples was used for prediction model

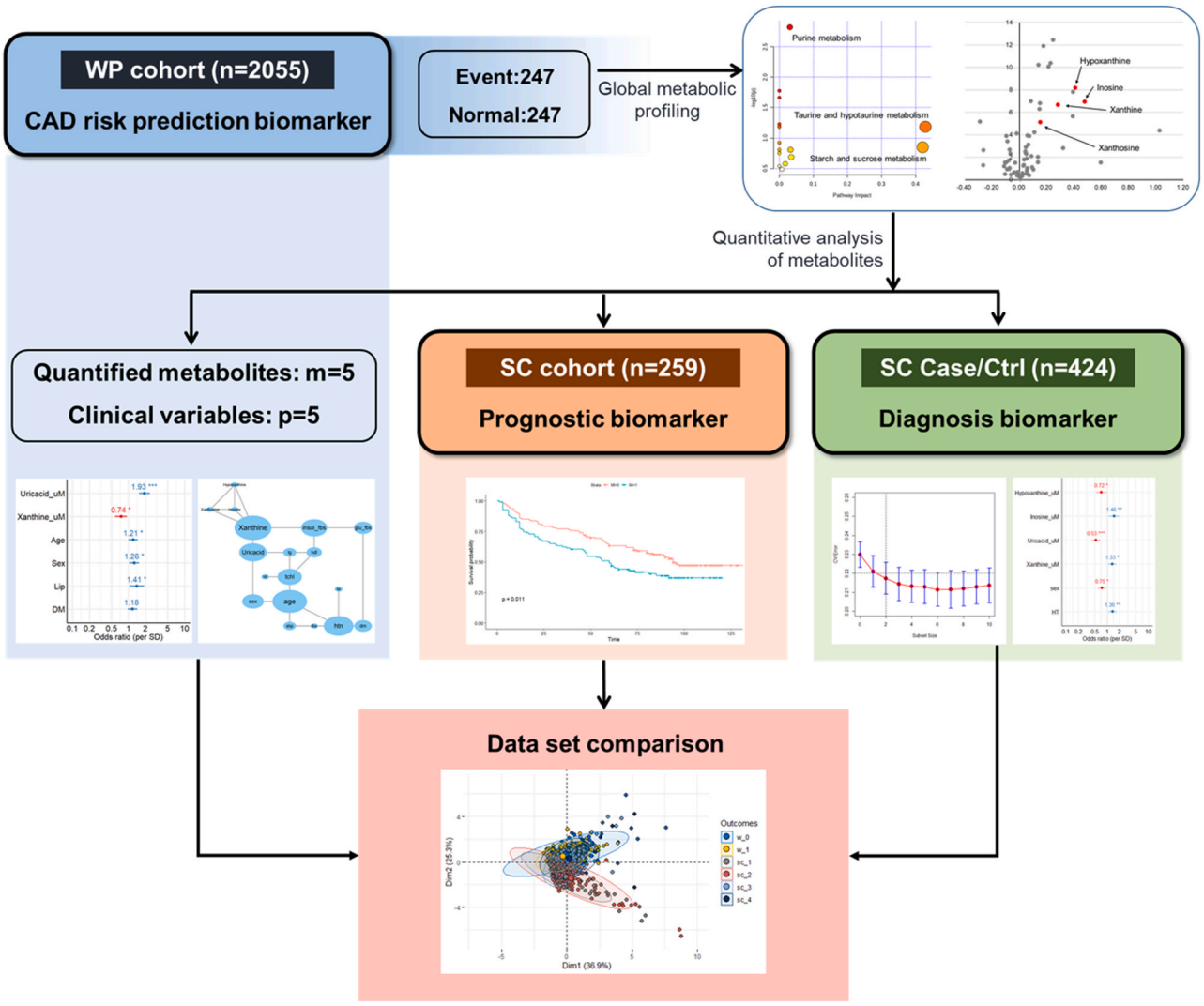

Fig. 1. Flow diagram describing the research process. In the Wonju and Pyeongchang (WP) cohort study, 2738 participants (247 people with CAD events within 10 years and 1814 people without CAD events within 10 years) were enrolled, and all the participants did not have CAD at baseline. Altered metabolic pathways were identified by global metabolic profiling using UPLC-Q/TOF MS. Purine metabolism was selected as the target pathway. The levels of 5 purine metabolites in participants' plasma were quantified by using UHPLC-TQ MS/MS. Descriptive analysis was performed for 5 metabolite abundances and 5 clinical variables. Logistic regression-based feature selection model development was performed for risk prediction model development by using 5 metabolic features and 5 clinical features from 247 event samples and 247 matched normal counterparts. Additionally, correlation-based network analysis was separately performed by using unmatched normal sample data. In the Shinchon (SC) cohort study, 259 participants (139 CAD patients with MACCEs within 10 years and 120 patients without MACCEs within 10 years) were enrolled, and targeted metabolite profiling was performed. Survival analysis was performed to test the difference in the time to relapse between the patients diagnosed with angina and the patients diagnosed with MI. Logistic regression-based machine learning model development was performed similar to the WP cohort to develop a secondary prevention model. Similarly, in the Shinchon (SC) case control study, 424 participants (150 patients with angina and 274 normal counterparts) were enrolled, and targeted metabolite quantification was performed. Logistic regression-based feature selection was also performed to develop a diagnostic model.

building by using WP cohort data. The best model and 1-sd rule model were evaluated based on the cross validation error (CVError) from logistic regression using 100 repeated 5-fold cross validations (Fig. 2A). The CVError of the model including only clinical data did not decrease with an increasing number of features. However, while the machine learning model selected features from 5 clinical features and 5 metabolic features together, the CVError instantly decreased with the inclusion of metabolic features, uric acid and xanthine. The best model with the smallest CVError was the logistic model with 6 features: sex, age, hyperlipidemia, diabetes, uric acid and xanthine. Considering the 1-sd rule to avoid overfitting issues, we could select the model only with hyperlipidemia, uric acid and xanthine. The estimated adjusted odds ratios (AORs) of features in our best model describe each metabolic feature's independent impact on CAD prevention (Fig. 2B). Elevated concentrations of uric acid and decreased xanthine levels were independently associated with future MACCE risk in the WP cohort group when age, sex, hyperlipidemia, and diabetes were adjusted.

Additionally, we evaluated the performance of the following three models: 1) the full model using all 10 features (5 metabolites and 5 clinical variables), 2) a model with only 5 metabolites, and 3) a model with only 5 clinical variables. Based on the mean are under the curve (AUC) from logistic regression by using 5-fold cross validations, the performance of each model was re-evaluated (Table 5). Similar to CVerror-based feature selection procedures, the model with age, sex, hyperlipidemia, diabetes, uric acid and xanthine performed better than the other models. Again, as the feature selection procedure results show in Fig. 2A, the mean AUC from cross validations of the model built with only clinical variables was lower than that of the model built with serum

Table 1
Baseline characteristics of WP cohort data.

	Non-MACCE (n = 1808)		MACCE (n = 247)		p-value
Sex (Female)	57.85		57.9		1
Age, years	54	[48,62]	59	[52,64]	< 0.0001
Hypertension	17.09		29.55		< 0.0001
Hyperlipidemia	1.60		3.64		0.0477
Diabetes	6.14		10.93		0.0072
Systolic blood pressure, mm Hg	130.0	[110.0140.0]	130.0	[120.0, 140.0]	0.0147
Diastolic blood pressure, mm Hg	80.0	[71.0, 90.0]	90.0	[80.0,90.0]	0.0089
Total cholesterol, mmol/L	123.0	[84.0, 187.0]	139.0	[97.5205.0]	0.0312
LDL-C, mmol/L	116.0	[96.0, 137.5]	121.0	[99.5, 140.5]	0.7423
HDL-C, mmol/L	46.0	[39.5, 54.0]	45.0	[38.0, 51.0]	0.0962
Insulin-FBS, mmol/L	7.20	[5.65, 9.05]	7.20	[5.60, 9.75]	0.7423
Glucose-FBS, mmol/L	92.0	[85.0, 99.0]	92.0	[85.0102.0]	0.6872

Values are either percentages or medians [25%, 75% quantile]. The p-values are from the Wilcoxon test.
LDL-C: low-density lipoprotein cholesterol, HDL-C: high-density lipoprotein cholesterol, insulin-FBS: fasting blood serum insulin level.

Table 2
Baseline characteristics of SC cohort data.

	Non-MACCE (n = 120)		MACCE (n = 139)		p-value
Sex (Female)	24.2		25.9		0.7498
Age, years	63.5	[56,70]	63	[55,69]	0.6023
Hypertension	59.2		61.9		0.6583
Hyperlipidemia	29.2		71.9		0.8449
Diabetes	33.3		38.1		0.4238
Currently smoking	25.8		20.1		0.2778

Values are either percentages or medians [25%, 75% quantile]. The p-values are from the Wilcoxon test.

Table 3
Baseline characteristics of SC case-control data.

	Normal (n = 274)		Stable angina (n = 150)		p-value
Sex (Female)	21.2		14.0		0.0703
Age, years	63	[55,70]	63	[56,69]	0.8264
Hypertension	52.9		69.3		0.0010
Hyperlipidemia	31.8		42.7		0.0250
Diabetes	32.8		40.0		0.1413

Values are either percentages or medians [25%, 75% quantile]. The p-values are from the Wilcoxon test.

Table 4
Concentration of serum metabolites in WP cohort data.

Metabolites	Non-MACCE (n = 1808)		MACCE (n = 247)		p-value
Uric acid (μM)	186.89	[149.69, 229.26]	210.34	[170.03, 258.74]	< 0.0001
Hypoxanthine (μM)	7.54	[3.78, 13.27]	6.29	[3.22, 11.33]	0.0096
Xanthine (μM)	1.10	[0.77, 1.64]	1.10	[0.74, 1.65]	0.9128
Inosine (μM)	0.20	[0.09, 0.36]	0.18	[0.08, 0.31]	0.0237
Xanthosine (nM)	24.40	[17.73, 34.62]	30.37	[149.69, 29.26]	0.0072

Values are medians [25%, 75% quantile]. The p-values are from the Wilcoxon test.

concentrations of purine metabolites (Table 5).

3.4. Correlation coefficient-based network analysis

To investigate the relationship among our clinical and metabolic features at the population level, unmatched non-MACCE group data were separately processed to build correlation coefficient-based network analysis (Fig. 3). Quantified purine metabolites and assessed clinical parameters are represented as nodes in the network graph. Interestingly, xanthine had an association with all four other metabolites, and its stress measure was larger than that of any other feature. Considering that xanthine was also selected to build the risk prediction model above, xanthine may work as a hub metabolite in purine metabolism that is correlated with other metabolites and clinical features. Furthermore, uric acid, another selected metabolic feature in our risk prediction model, was also highlighted, as shown in Fig. 3. Therefore, our concordant results demonstrated that xanthine and uric acid are representative features among our 10 variables, and network analysis results cautiously provided the reason why xanthine and uric acid should be included in the CAD risk prediction model.

3.5. Applying selected features to predict the prognosis of stable CAD and MI patients

To examine whether serum purine metabolites can be applicable for the development of a secondary prevention model for CAD, equivalent targeted metabolomics analysis and machine learning analysis were conducted. At this time, the research was performed by using an SC cohort study, which included 259 patients with stable CAD ($n = 106$) or myocardial infarction ($n = 161$). The MACCE group refers to a group of patients with MACCEs during cohort tracking, and the control group refers to its corresponding counterpart.

The association between previous MI history and MACCEs was tested by using the log rank test of Kaplan-Meier survival curves (p-value: 0.011, Fig. 4) [25]. Patients who were previously diagnosed with MI had distinctly earlier MACCEs than patients without an MI diagnosis. If two groups with different disease histories had been merged for the analysis, disease history could have been a confounding factor in our statistical learning process. For this reason, we performed survival analysis via the Cox proportional hazard model on data from all patients, patients with MI history, and patients with stable CAD history only separately. Although the p-value from the estimated hazard ratio of elevated xanthine levels on MACCE relapse was smaller than the p-values from other metabolites, all the estimated hazard ratios were not statistically significant (Supplemental Table 1). Similar to the WP cohort, logistic regression was also carried out by 5-fold cross validation to build a model for secondary prevention of CAD patients. Based on the average AUCs of the statistical analysis, purine metabolites could be applicable for the prediction of future MACCEs in stable MI patients, but they could be insufficient for secondary prevention in stable CAD patients. Regardless of the performance of the best models from each group, intriguingly, xanthine was again included in the best set of features for both groups (Table 6).

3.6. Applying selected features for MI diagnosis model building

Next, we wanted to determine the following: 1) whether serum purine metabolite concentration could be applicable to stable CAD diagnosis and 2) whether the same set of features are selected for diagnosis model development, similar to our primary/secondary prevention models. At this time, SC case-control data were also equivalently processed to clarify our research questions. Here, we could develop a diagnosis model with acceptable performance. The best performing model was built with 6 features: sex, hypertension, uric acid, hypoxanthine, xanthine, and inosine (Fig. 5, Table 7). Interestingly, xanthine was again selected as a metabolic feature for the stable CAD diagnosis

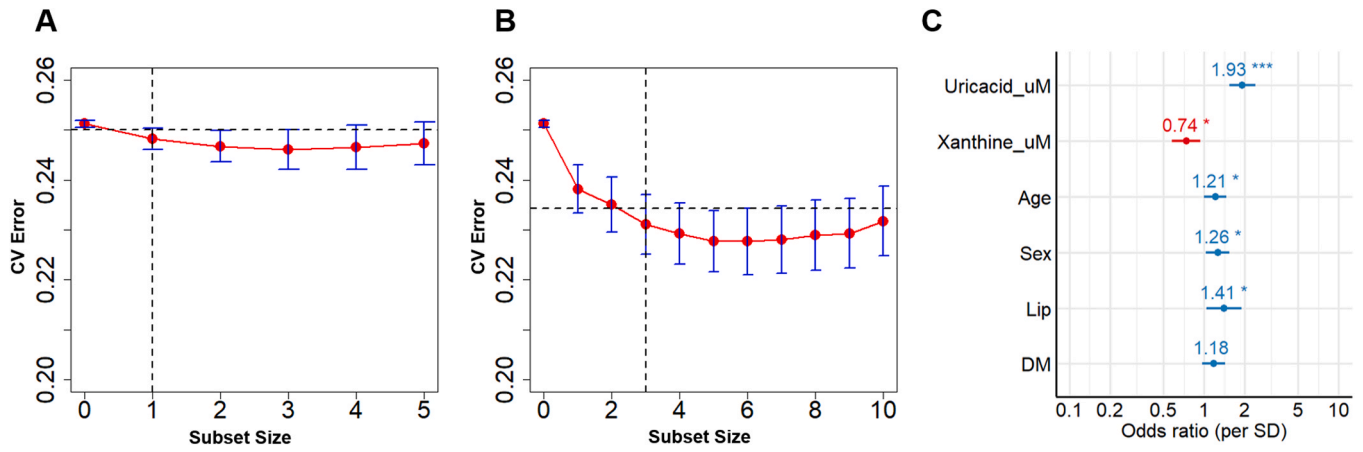

Fig. 2. Building model for primary prevention by using logistic regression (5-fold cross validation). Cross validation (CV) error plot of the logistic regression by 100 repeated 5-fold cross validations. (**A**) using only 5 clinical features. (**B**) Using all 5 clinical features and 5 metabolic features. (C) OR (95% CI) of CAD per SD of features in the best model (with the lowest CV error) derived from the WP cohort study.

Table 5
Machine learning models for the primary prevention of CAD (WP cohort).

	MACCE risk prediction models	Mean AUC
Full model	Y ~ Age + Sex + HTN + Lip + DM + Uric acid_uM + Hypoxanthine_uM + Xanthine_uM + Inosine_uM + Xanthosine_nM	0.68
5 metabolites	Y ~ Uric acid_uM + Hypoxanthine_uM + Xanthine_uM + Inosine_uM + Xanthosine_nM	0.63
Best model	Y ~ Age + Sex + HTN + Lip + DM + Uric acid_uM + Xanthine_uM	0.70
1-sd model	Y ~ Lip + Uric acid_uM + Xanthine_uM	0.67
Only clinical	Y ~ Age + Sex + HTN + Lip + DM	0.59

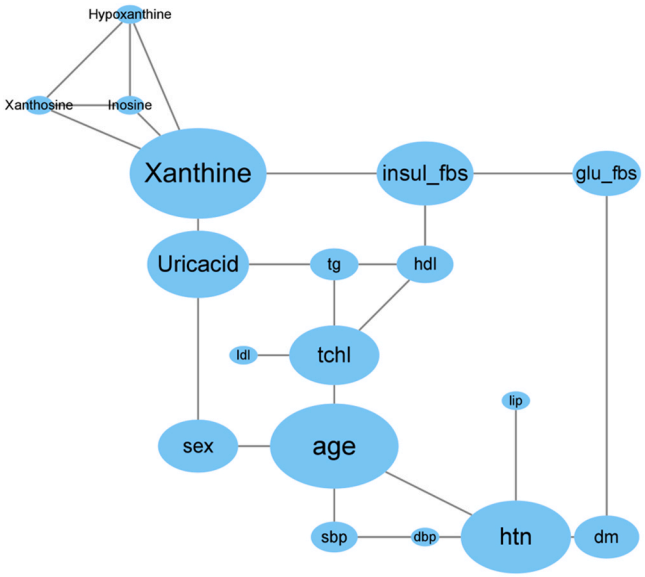

Fig. 3. Correlation coefficient-based network analysis result. Network visualization of correlation-based relationships among 5 purine-related metabolites and 13 clinical parameters. The width and height of nodes were scaled using stress centrality measures.

model, similar to the other two datasets. We cautiously assume that serum xanthine concentration is one of the key biomarkers representing altered purine metabolism in CAD patients. However, the directional effects of serum xanthine concentration for outcome measurement were not concordant among our three datasets. In the WP cohort data, a lower serum xanthine level in a person tended to be associated with a higher probability of having a MACCE from normal baseline within 10 years. However, an increase in serum xanthine levels was associated with the recurrence of a MACCE in 11 years or the diagnosis of stable CAD. In addition, although elevated serum uric acid levels are known to be CAD biomarkers [26], a decrease in the serum uric acid concentration was highly correlated with stable CAD diagnosis in our data. Considering the consistency in the analytical platform and sufficiently low coefficient of variance values from our LC-MS data (Supplemental Fig. 2), this may be due to two reasons: 1) different roles of circulating purine metabolites in primary vs secondary prevention and diagnosis of CAD or 2) changes in their mutual interaction or interaction with clinical features.

3.7. PCA of three targeted metabolomics datasets applied to prediction model development

To clarify the differences in our three targeted metabolomics datasets, we conducted principal component analysis (PCA) (Fig. 6). As shown in the PCA plot, the gray colored non-MACCE group from the SC cohort and the red colored counterpart from the same SC cohort were different from the other groups in either the WP cohort data or SC case control data. The PCA plot discrepancy could be attributed to the difference in the medical history of the participants. The SC cohort study recruited participants with MACCEs before enrollment, but participants in the other two studies were either normal participants or patients with stable CAD at baseline.

4. Discussion

Most importantly, this is the first study to systemically compare machine learning models derived from the same set of metabolic features for primary/secondary prevention and diagnosis of a certain disease. Since it is technically challenging to generate and quantify comparable metabolic datasets, there have been no attempts to compare different characteristics of metabolic features in the prevention or diagnosis of a disease.

Our work described purine metabolism as a metabolic determinant of CAD risk prediction via metabolic profiling based on LC-MS and

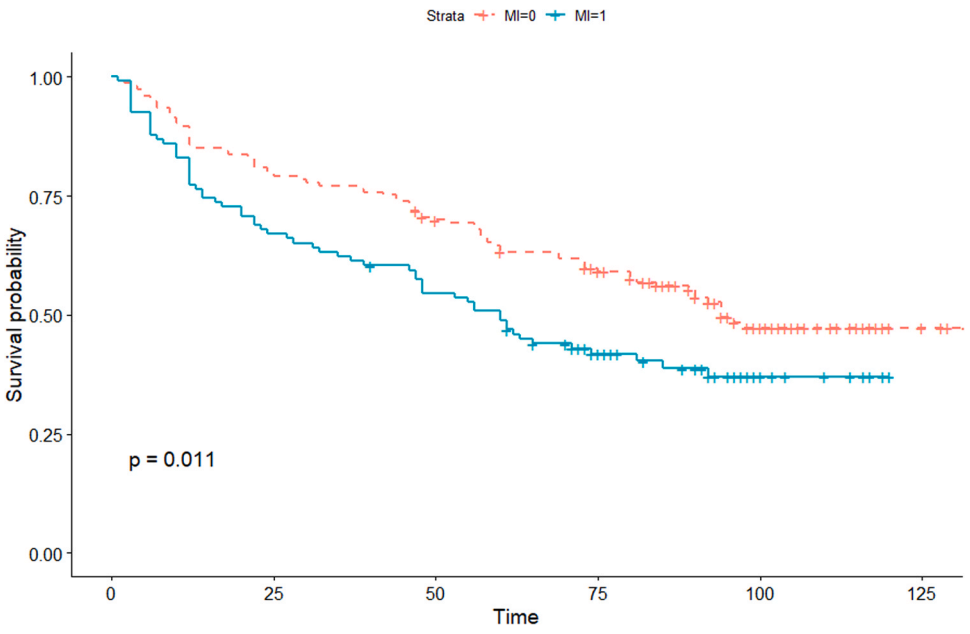

Fig. 4. Survival plot of patients previously with MI ($n = 106$) vs without MI (stable CAD, $n = 153$). The probability without MACCEs in the cohort was analyzed by using the log-rank test in a survival plot (*P*-value: 0.011). The blue line refers to patients previously having MI, and the red dotted line refers to patients without MI. Two groups showed significantly different times to MACCE event results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Machine learning model for the secondary prevention of CAD patients (SC cohort).

History	Best model	Mean AUC
Angina	Y ~ Xanthine_uM	0.58
MI	Y ~ age + DM + Xanthine_uM	0.71

revisited the association between metabolite features from altered purine metabolism and CAD. We also applied a targeted MS analysis platform to quantify the serum concentrations of 5 purine metabolites in 2738 blood samples from three different large datasets. The WP cohort, SC cohort, and SC case control datasets were analyzed for first/secondary prevention and diagnosis of CAD. These datasets were equivalently processed to propose prediction models and further provided the idea that inconsistent metabolic features could be selected for different stages of disease prevention or diagnosis. The results from our research support a potential role of purine metabolism-related metabolic features in predicting 1. the risk of disease at the normal baseline, 2. relapse of a MACCE in patients with MI history, and 3. the possibility of stable CAD. In contrast, our research simultaneously proposed the limitation of purine-related metabolic features on developing machine learning models for secondary prevention in stable CAD patients, who have relatively mild symptoms.

Although previous studies have introduced machine-learning algorithms for the diagnosis of CAD [10,11,27], we first demonstrated the role of serum metabolites as features in the models for the diagnosis and

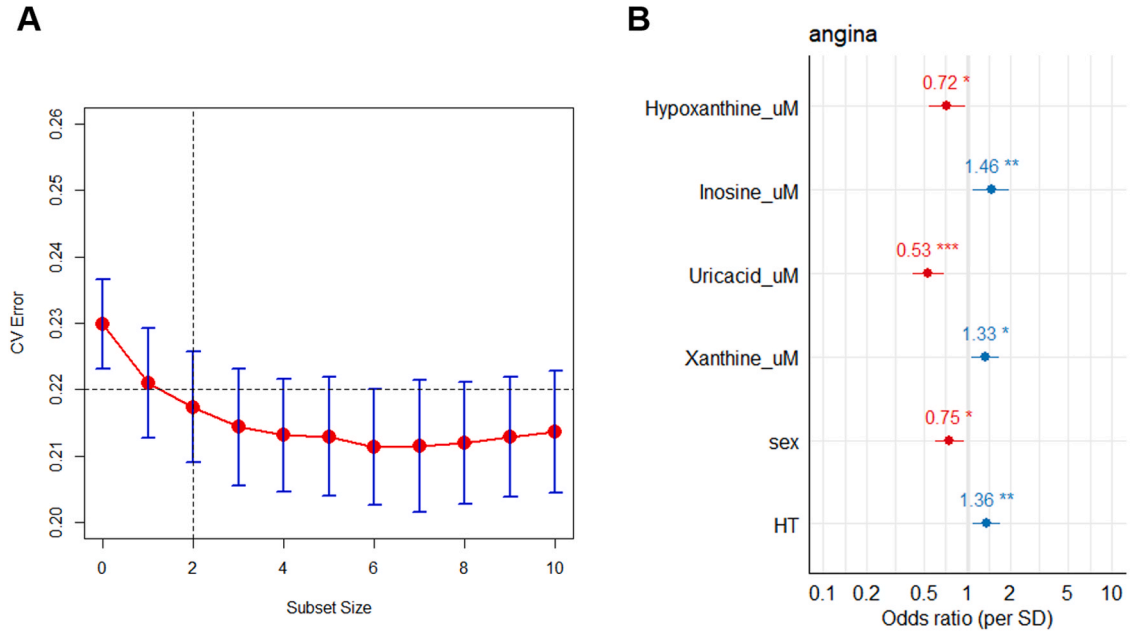

Fig. 5. Building model for angina diagnosis by using logistic regression (5-fold cross validation). (**A**) Cross validation (CV) error plot of the logistic regression by 100 repeated 5-fold cross validations. using all 5 clinical features and 5 metabolic features. (**B**) OR (95% CI) of CAD per SD of features in the best model (with the lowest CV error) derived from the SC case-control study.

Table 7
Stable CAD diagnosis model development results (SC case-control).

	Angina diagnosis models	Mean AUC
Full model	Y ~ Age + Sex + HTN + Lip + DM + Uric acid_uM + Hypoxanthine_uM + Xanthine_uM + Inosine_uM + Xanthosine_nM	0.68
5 metabolites	Y ~ Uric acid_uM + Hypoxanthine_uM + Xanthine_uM + Inosine_uM + Xanthosine_nM	0.63
Best model	Y ~ Sex + HTN + Uric acid_uM + Hypoxanthine_uM + Xanthine_uM + Inosine_uM	0.70
1-sd model	Y ~ HTN + Uric acid_uM	0.67
Only clinical	Y ~ Age + Sex + HTN + Lip + DM	0.59

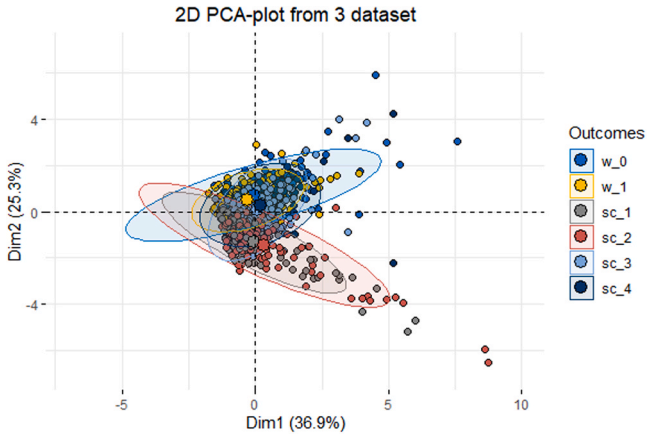

Fig. 6. PCA result from 3 datasets. w_0: non-MACCE group in the WP cohort study ($n = 247$), w_1: MACCE group in the WP cohort study ($n = 247$), sc_1: non-MACCE group in the SC cohort study ($n = 120$), sc_2: MACCE group in the SC cohort study ($n = 139$), sc_3: non-MACCE group in the SC case control study ($n = 274$), sc_4: MACCE group in the SC case control study ($n = 150$). (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

primary/secondary prevention of CAD based on large-sized population-level cohort data. Recently, deep learning methods have often been applied to the analysis of angiography or electrocardiogram (ECG) signals for the diagnosis of CAD [27–32]. However, medical decisions based on cardio images or signals limit the decision about the diagnosis of the current medical status. In contrast, metabolomics has been considered an emerging technology in the prevention of cancer or other metabolic diseases, including CAD [33,34]. However, due to the scarcity of metabolomics from large-sized, population-based cohort data, there is still extensive research to be performed.

Over a decade, our knowledge of purine metabolism disorders in CAD continues to grow. It is known that ubiquitous xanthine oxidase (XO) controls the rate-limiting step of purine catabolism [35,36]. Since XO transfers electrons from hypoxanthine or xanthine to oxygen and NAD+, the generated powerful oxidant could cause cardiac tissue damage [37]. Defects in XO are known to be associated with type I xanthinuria and blood pressure changes [38,39]. Researchers suggest that hypoxia due to ischemia results in enhanced ATP catabolism and increases the hypoxanthine level [40]. Reperfusion after ischemia could cause ischemia-reperfusion injury by XO-generated ROS [41]. Interestingly, xanthine was consistently selected as a major metabolic determinant for our prediction models from three different datasets. Turgan et al. demonstrated an association between elevated urinary xanthine levels and acute coronary syndromes. In addition, they demonstrated that xanthine was negatively correlated with uric acid only in patients but not in healthy controls [42]. Intriguingly, although we could not observe a negative correlation between the levels of xanthine and uric acid in blood, their directional effects for CAD were always opposite in our three different datasets. In fact, the association between serum XO or uric acid levels and CAD has often been investigated in prior studies, but serum xanthine levels were first demonstrated as a notable biomarker for CAD in our research.

Furthermore, uric acid, the ultimate product of purine catabolism, was also selected as a feature in models for CAD risk prediction and diagnosis of stable CAD. Uric acid is known to be a strong antioxidant [43,44], but elevated serum uric acid is associated with CAD and its known risk factors, such as hypertension, renal disease, metabolic syndrome, and oxidative stress [26]. In our results, we identified a positive correlation between serum uric acid and CAD risk in a WP cohort study when serum xanthine level and 4 clinical variables, age, sex, hyperlipidemia, and diabetes, were adjusted in the model. However, in the SC case-control study, a decrease in serum uric acid was significantly associated with stable CAD diagnosis when hypoxanthine, inosine, xanthine and 2 clinical variables, sex and hypertension, were included in the prediction model. Since it is debated that serum uric acid level is an independent factor of CAD [45], the observed negative correlation between uric acid and stable angina diagnosis in our machine learning model could be affected by the inclusion of hypertension or other metabolite features. Importantly, as a feature selection procedure was conducted in our research, the most informative set of features was identified for model development, and our approach tentatively avoids referring to an indirect factor as a determinant of multifactor diseases, such as CAD.

In addition, in the case of patients already diagnosed with CAD or related metabolic disease, their medication is to be reported during the data acquisition step to avoid distorting the role of features in the pathological environment [46,47]. Possible associations between the metabolic biomarkers and phenotypes could be confounded by the prescribed drugs for CAD or other metabolic diseases, such as kidney disease, hypertension or diabetes. Therefore, our proposed estimation of the adjusted odds ratio from multiple regression could be more applicable to CAD patient care than the simple odds of each feature.

In fact, our machine-learning results highlighted the possible opposing effects of metabolic features on CAD outcomes at different disease prevention phases. It is worth noting that studies have already reported the opposing effects of clinical parameters and microRNA features in the primary and secondary prevention of CAD [48,49]. Circulating MIR-197 and MIR-223 were reported to be associated with MI, and their reported HR directions for primary and secondary prevention were not consistent. HRs in primary MI prevention were estimated as less than one, but those in secondary prevention were estimated as greater than one. Similarly, Xia et al. reported the inconsistent impact of clinical parameters (sex) on primary and secondary CVD prevention [50]. Therefore, we suggest that it is plausible to find the same features with opposing effects when machine-learning algorithms estimate prediction models for different prevention phases. This evidence suggests that biomarker discovery should be conducted by using appropriate datasets for the aim of a study. For example, when we explore the prediction biomarker for disease risk, the biomarker from a case-control study is inadequate for research purposes.

5. Conclusions

We equivalently processed three large datasets in a controlled manner and found that purine metabolism is associated with CAD disease risk and status. In particular, xanthine was a determinant metabolic

feature for the developed machine learning models in our cohort datasets and for that in the case control data set. This study is the first attempt to systemically compare the roles of metabolic biomarkers in three datasets with different research aims. Since metabolite quantification can be easily biased by experimental procedures, establishing a high level of reproducibility is analytically challenging in this type of research using a couple of large datasets. For this reason, our initial research first provides evidence that xanthine and uric acid show opposing effects in machine learning models for primary and secondary CAD prevention. Furthermore, potentially the same research strategy could be also valid for other biomarker studies. Our results can extend our knowledge to appreciate possible changes in the role of metabolic features in machine learning model developments for primary/secondary prevention or diagnosis of other diseases.

Funding source

This work was supported by the Korea Basic Science Institute (C170200); the National Research Foundation of Korea (NRF) grant funded by the Korea government (Nos. 2019M3A9D5A01102796 and 2020R1A2C2007835).

CRediT authorship contribution statement

All authors participated in the research design. SJ performed metabolomics analysis. EA performed statistical analysis and wrote the first version of the manuscript. SBK and SHL recruited participants and collected the samples. GSH supervised the research. All authors revised the draft and finalized the manuscript.

Conflict of interest statement

None.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopha.2021.111621.

References

[1] WHO, Cardiovascular diseases (CVDs), (2017). ⟨https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)⟩ (Accessed November 13, 2020).
[2] H. Kim, S. Kim, S. Han, P.P. Rane, K.M. Fox, Y. Qian, H.S. Suh, Prevalence and incidence of atherosclerotic cardiovascular disease and its risk factors in Korea: a nationwide population-based study, BMC Public Health 19 (2019) 1112, https://doi.org/10.1186/s12889-019-7439-0.
[3] M. Ruiz-Canela, A. Hruby, C.B. Clish, L. Liang, M.A. Martínez-González, F.B. Hu, Comprehensive metabolomic profiling and incident cardiovascular disease: a systematic review, J. Am. Heart Assoc. 6 (2017), https://doi.org/10.1161/JAHA.117.005705.
[4] U.A. Ajani, E.S. Ford, Has the risk for coronary heart disease changed among U.S. adults? J. Am. Coll. Cardiol. 48 (2006) 1177–1182, https://doi.org/10.1016/j.jacc.2006.05.055.
[5] V.J. Dzau, E.M. Antman, H.R. Black, D.L. Hayes, J.E. Manson, J. Plutzky, J. J. Popma, W. Stevenson, The cardiovascular disease continuum validated: clinical evidence of improved patient outcomes: part II: clinical trial evidence (acute coronary syndromes through renal disease) and future directions. Circulation 114 (2006) 2871–2891, https://doi.org/10.1161/CIRCULATIONAHA.106.655761.
[6] J. Lee, Y. Jung, J.Y. Park, S.-H. Lee, D.H. Ryu, G.-S. Hwang, LC/MS-based polar metabolite profiling reveals gender differences in serum from patients with myocardial infarction, J. Pharm. Biomed. Anal. 115 (2015) 475–486, https://doi.org/10.1016/j.jpba.2015.08.009.
[7] E.J. Benjamin, P. Muntner, A. Alonso, M.S. Bittencourt, C.W. Callaway, A. P. Carson, A.M. Chamberlain, A.R. Chang, S. Cheng, S.R. Das, F.N. Delling, L. Djousse, M.S.V. Elkind, J.F. Ferguson, M. Fornage, L.C. Jordan, S.S. Khan, B. M. Kissela, K.L. Knutson, T.W. Kwan, D.T. Lackland, T.T. Lewis, J.H. Lichtman, C. T. Longenecker, M.S. Loop, P.L. Lutsey, S.S. Martin, K. Matsushita, A.E. Moran, M. E. Mussolino, M. O'Flaherty, A. Pandey, A.M. Perak, W.D. Rosamond, G.A. Roth, U. K.A. Sampson, G.M. Satou, E.B. Schroeder, S.H. Shah, N.L. Spartano, A. Stokes, D. L. Tirschwell, C.W. Tsao, M.P. Turakhia, L.B. VanWagner, J.T. Wilkins, S.S. Wong, S.S. Virani, Heart disease and stroke statistics-2019 update: a report from the american heart association, Circulation 139 (2019) e56–e528, https://doi.org/10.1161/CIR.0000000000000659.
[8] M.S. Sabatine, E. Liu, D.A. Morrow, E. Heller, R. McCarroll, R. Wiegand, G. F. Berriz, F.P. Roth, R.E. Gerszten, Metabolomic identification of novel biomarkers of myocardial ischemia, Circulation 112 (2005) 3868–3875, https://doi.org/10.1161/CIRCULATIONAHA.105.569137.
[9] J. Teul, F.J. Rupérez, A. Garcia, J. Vaysse, S. Balayssac, V. Gilard, M. Malet-Martino, J.L. Martin-Ventura, L.M. Blanco-Colio, J. Tuñón, J. Egido, C. Barbas, Improving metabolite knowledge in stable atherosclerosis patients by association and correlation of GC-MS and ^{1}H NMR fingerprints, J. Proteome Res. 8 (2009) 5580–5589, https://doi.org/10.1021/pr900668v.
[10] G.D. Lewis, R. Wei, E. Liu, E. Yang, X. Shi, M. Martinovic, L. Farrell, A. Asnani, M. Cyrille, A. Ramanathan, O. Shaham, G. Berriz, P.A. Lowry, I.F. Palacios, M. Taşan, F.P. Roth, J. Min, C. Baumgartner, H. Keshishian, T. Addona, V. K. Mootha, A. Rosenzweig, S.A. Carr, M.A. Fifer, M.S. Sabatine, R.E. Gerszten, Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury, J. Clin. Investig. 118 (2008) 3503–3512, https://doi.org/10.1172/JCI35111.
[11] Y. Fan, Y. Li, Y. Chen, Y.-J. Zhao, L.-W. Liu, J. Li, S.-L. Wang, R.N. Alolga, Y. Yin, X.-M. Wang, D.-S. Zhao, J.-H. Shen, F.-Q. Meng, X. Zhou, H. Xu, G.-P. He, M.-D. Lai, P. Li, W. Zhu, L.-W. Qi, Comprehensive metabolomic characterization of coronary artery diseases, J. Am. Coll. Cardiol. 68 (2016) 1281–1293, https://doi.org/10.1016/j.jacc.2016.06.044.
[12] M. Ruiz-Canela, E. Toledo, C.B. Clish, A. Hruby, L. Liang, J. Salas-Salvadó, C. Razquin, D. Corella, R. Estruch, E. Ros, M. Fitó, E. Gómez-Gracia, F. Arós, M. Fiol, J. Lapetra, L. Serra-Majem, M.A. Martínez-González, F.B. Hu, Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial, Clin. Chem. 62 (2016) 582–592, https://doi.org/10.1373/clinchem.2015.251710.
[13] S.M. Haffner, S. Lehto, T. Rönnemaa, K. Pyörälä, M. Laakso, Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction, N. Engl. J. Med. 339 (1998) 229–234, https://doi.org/10.1056/NEJM199807233390404.
[14] W.B. Kannel, J.T. Doyle, R.J. Shephard, J. Stamler, P.S. Vokonas, Prevention of cardiovascular disease in the elderly, J. Am. Coll. Cardiol. 10 (1987) 25A–28A, https://doi.org/10.1016/s0735-1097(87)80443-6.
[15] S. Kume, S. Araki, N. Ono, A. Shinhara, T. Muramatsu, H. Araki, K. Isshiki, K. Nakamura, H. Miyano, D. Koya, M. Haneda, S. Ugi, H. Kawai, A. Kashiwagi, T. Uzu, H. Maegawa, Predictive properties of plasma amino acid profile for cardiovascular disease in patients with type 2 diabetes, PLoS One 9 (2014), 101219, https://doi.org/10.1371/journal.pone.0101219.
[16] S. Kiechl, J. Willeit, In a nutshell: findings from the Bruneck Study, Gerontology 65 (2019) 9–19, https://doi.org/10.1159/000492329.
[17] C. Stegemann, R. Pechlaner, P. Willeit, S.R. Langley, M. Mangino, U. Mayr, C. Menni, A. Moayyeri, P. Santer, G. Rungger, T.D. Spector, J. Willeit, S. Kiechl, M. Mayr, Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study, Circulation 129 (2014) 1821–1831, https://doi.org/10.1161/CIRCULATIONAHA.113.002500.
[18] S. Kiechl, E. Lorenz, M. Reindl, C.J. Wiedermann, F. Oberhollenzer, E. Bonora, J. Willeit, D.A. Schwartz, Toll-like receptor 4 polymorphisms and atherogenesis. N. Engl. J. Med. 347 (2002) 185–192, https://doi.org/10.1056/NEJMoa012673.
[19] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D.S. Wishart, J. Xia, MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis, Nucleic Acids Res. 46 (2018) W486–W494, https://doi.org/10.1093/nar/gky310.
[20] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30, https://doi.org/10.1093/nar/28.1.27.
[21] E. Saccenti, M. Suarez-Diez, C. Luchinat, C. Santucci, L. Tenori, Probabilistic networks of blood metabolites in healthy subjects as indicators of latent cardiovascular risk, J. Proteome Res. 14 (2015) 1101–1111, https://doi.org/10.1021/pr501075r.
[22] R Core Team, R: A language and environment for statistical computing, (2020). https://www.r-project.org/.
[23] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.
[24] U. Brandes, T. Erlebach. G. für Informatik, G. fèur Informatik, Network Analysis: Methodological Foundations, Springer, 2005. https://books.google.co.kr/books?id=VIMSPCIafakC.
[25] J.M. Bland, D.G. Altman, The logrank test, BMJ 328 (2004) 1073, https://doi.org/10.1136/bmj.328.7447.1073.
[26] M. Kanbay, M. Segal, B. Afsar, D.-H. Kang, B. Rodriguez-Iturbe, R.J. Johnson, The role of uric acid in the pathogenesis of human cardiovascular disease, Heart 99 (2013) 759–766, https://doi.org/10.1136/heartjnl-2012-302535.
[27] R. Alizadehsani, M. Abdar, M. Roshanzamir, A. Khosravi, P.M. Kebria, F. Khozeimeh, S. Nahavandi, N. Sarrafzadegan, U.R. Acharya, Machine learning-based coronary artery disease diagnosis: a comprehensive review, Comput. Biol. Med. 111 (2019), 103346, https://doi.org/10.1016/j.compbiomed.2019.103346.
[28] Y. Shen, Z. Fang, Y. Gao, N. Xiong, C. Zhong, X. Tang, Coronary arteries segmentation based on 3D FCN with attention gate and level set function, IEEE Access 7 (2019) 42826–42835, https://doi.org/10.1109/ACCESS.2019.2908039.
[29] U.R. Acharya, H. Fujita, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals, Inf. Sci. 415–416 (2017) 190–198, https://doi.org/10.1016/j.ins.2017.06.027.

[30] J. Betancur, F. Commandeur, M. Motlagh, T. Sharir, A.J. Einstein, S. Bokhari, M. B. Fish, T.D. Ruddy, P. Kaufmann, A.J. Sinusas, E.J. Miller, T.M. Bateman, S. Dorbala, M. Di Carli, G. Germano, Y. Otaki, B.K. Tamarappoo, D. Dey, D. S. Berman, P.J. Slomka, Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: a multicenter study, JACC Cardiovasc. Imaging 11 (2018) 1654–1663, https://doi.org/10.1016/j.jcmg.2018.01.020.

[31] J.H. Tan, Y. Hagiwara, W. Pang, I. Lim, S.L. Oh, M. Adam, R.S. Tan, M. Chen, U. R. Acharya, Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals, Comput. Biol. Med. 94 (2018) 19–26, https://doi.org/10.1016/j.compbiomed.2017.12.023.

[32] U.R. Acharya, H. Fujita, O.S. Lih, M. Adam, J.H. Tan, C.K. Chua, Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network, Knowl. Based Syst. 132 (2017) 62–71, https://doi.org/10.1016/j.knosys.2017.06.003.

[33] A.C. Dona, S. Coffey, G. Figtree, Translational and emerging clinical applications of metabolomics in cardiovascular disease diagnosis and treatment, Eur. J. Prev. Cardiol. 23 (2016) 1578–1589, https://doi.org/10.1177/2047487316645469.

[34] Y.S. Kim, P. Maruvada, J.A. Milner, Metabolomics in biomarker discovery: future uses for cancer prevention, Future Oncol. 4 (2008) 93–102, https://doi.org/10.2217/14796694.4.1.93.

[35] D.A. Parks, D.N. Granger, Xanthine oxidase: biochemistry, distribution and physiology, Acta Physiol. Scand. Suppl. 548 (1986) 87–99.

[36] A. Meneshian, G.B. Bulkley, The physiology of endothelial xanthine oxidase: from urate catabolism to reperfusion injury to inflammatory signal transduction, Microcirculation 9 (2002) 161–175, https://doi.org/10.1038/sj.mn.7800136.

[37] C.E. Berry, J.M. Hare, Xanthine oxidoreductase and cardiovascular disease: molecular mechanisms and pathophysiological implications, J. Physiol. 555 (2004) 589–606, https://doi.org/10.1113/jphysiol.2003.055913.

[38] K. Ichida, Y. Amaya, K. Okamoto, T. Nishino, Mutations associated with functional disorder of xanthine oxidoreductase and hereditary xanthinuria in humans, Int. J. Mol. Sci. 13 (2012) 15475–15495, https://doi.org/10.3390/ijms131115475.

[39] L.E.J.M. Scheepers, F.-F. Wei, K. Stolarz-Skrzypek, S. Malyutina, V. Tikhonoff, L. Thijs, E. Salvi, C. Barlassina, J. Filipovský, E. Casiglia, Y. Nikitin, K. Kawecka-Jaszcz, P. Manunta, D. Cusi, A. Boonen, J.A. Staessen, I.C.W. Arts, Xanthine oxidase gene variants and their association with blood pressure and incident hypertension: a population study. J. Hypertens. 34 (2016) 2147–2154, https://doi.org/10.1097/HJH.0000000000001077.

[40] D.N. Granger, G. Rutili, J.M. McCord, Superoxide radicals in feline intestinal ischemia, Gastroenterology 81 (1981) 22–29.

[41] J.M. McCord, Oxygen-derived free radicals in postischemic tissue injury. N. Engl. J. Med. 312 (1985) 159–163, https://doi.org/10.1056/NEJM198501173120305.

[42] N. Turgan, S. Habif, C. Gülter, B. Şenol, I. Mutaf, D. Özmen, O. Bayindir, B. Boydak, N. Turgan, Urinary hypoxanthine and xanthine levels in acute coronary syndromes, Int. J. Clin. Lab. Res. 29 (1999) 162–165, https://doi.org/10.1007/s005990050084.

[43] Y.Y. Sautin, R.J. Johnson, Uric acid: the oxidant-antioxidant paradox, Nucleosides Nucleotides Nucleic Acids 27 (2008) 608–619, https://doi.org/10.1080/15257770802138558.

[44] B.N. Ames, R. Cathcart, E. Schwiers, P. Hochstein, Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis, Proc. Natl. Acad. Sci. U. S. A. 78 (1981) 6858–6862, https://doi.org/10.1073/pnas.78.11.6858.

[45] K.E.J.K.J.W.R.S.W.S.H.S.P.C.G. Lim Hong Euy Kim Seong Hwan, Clinical Value of Serum Uric Acid in Patients with Suspected Coronary Artery Disease, Korean J. Intern Med 25 (2010) 21–26, https://doi.org/10.3904/kjim.2010.25.1.21.

[46] K. Chan, X. Pu, P. Sandesara, R.N. Poston, I.A. Simpson, A.A. Quyyumi, S. Ye, R. S. Patel, Genetic variation at the ADAMTS7 locus is associated with reduced severity of coronary artery disease, J. Am. Heart Assoc. 6 (2017), https://doi.org/10.1161/jaha.117.006928.

[47] R.S. Patel, A.F. Schmidt, V, Tragante, R.O. McCubrey, M.V. Holmes, et al., Association of chromosome 9p21 with subsequent coronary heart disease events, Circ. Genomic Precis. Med. 12 (2019), https://doi.org/10.1161/CIRCGEN.119.002471.

[48] A. Zampetaki, P. Willeit, L. Tilling, I. Drozdov, M. Prokopi, J.-M. Renard, A. Mayr, S. Weger, G. Schett, A. Shah, C.M. Boulanger, J. Willeit, P.J. Chowienczyk, S. Kiechl, M. Mayr, Prospective study on circulating microRNAs and risk of myocardial infarction, J. Am. Coll. Cardiol. 60 (2012) 290–299, https://doi.org/10.1016/j.jacc.2012.03.056.

[49] C. Schulte, S. Molz, S. Appelbaum, M. Karakas, F. Ojeda, D.M. Lau, T. Hartmann, K. J. Lackner, D. Westermann, R.B. Schnabel, S. Blankenberg, T. Zeller, miRNA-197 and miRNA-223 predict cardiovascular death in a cohort of patients with symptomatic coronary artery disease, PLoS One 10 (2015), e0145930, https://doi.org/10.1371/journal.pone.0145930.

[50] X. Shijun, D. Xin, G. Lizhu, D. Jing, A. Clare, L.C. S.P, H.M. D, A. Hisatomi, Y. Yiqiang, Z. Yang, W. Shulin, G. Xuefeng, Z. Xianhui, L. Hongbo, C. Xiaoshu, A. C. S, D. Jianzeng, M. Changsheng, Sex differences in primary and secondary prevention of cardiovascular disease in China, Circulation 141 (2020) 530–539, https://doi.org/10.1161/CIRCULATIONAHA.119.043731.