Original Article

Novel network biomarkers profile based coronary artery disease risk stratification in Asian Indians

Rajani Kanth Vangala[1,2], Vandana Ravindran[2], Karthik Kamath[1], Veena S Rao[1], Hebbagodi Sridhara[2]

[1]Department of Tata Proteomics and Coagulation, [2]Elizabeth and Emmanuel Kaye Bioinformatics and Biostatistics Department, Thrombosis Research Institute, Narayana Hrudayalaya Hospital, Bangalore, Karnataka, India

Abstract

Background: Multi-marker approaches for risk prediction in coronary artery disease (CAD) have been inconsistent due to biased selection of specific know biomarkers. We have assessed the global proteome of CAD-affected and unaffected subjects, and developed a pathway network model for elucidating the mechanism and risk prediction for CAD.

Materials and Methods: A total of 252 samples (112 CAD-affected without family history and 140 true controls) were analyzed by Surface-Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS) by using CM10 cationic chips and bioinformatics tools.

Results: Out of 36 significant peaks in SELDI-TOF MS, nine peaks could do better discrimination of CAD subjects and controls (area under the curve (AUC) of 0.963) based on the Support Vector Machine (SVM) feature selection method. Of the nine peaks used in the model for discrimination of CAD-affected and unaffected, the *m/z* corresponding to 22,859 was identified as stress-related protein HSP27 and was shown to be highly associated with CAD (odds ratio of 3.47). The 36 biomarker peaks were identified and a network profile was constructed showing the functional association between different pathways in CAD.

Conclusion: Based on our data, proteome profiling with SELDI-TOF MS and SVM feature selection methods can be used for novel network biomarker discovery and risk stratification in CAD. The functional associations of the identified novel biomarkers suggest that they play an important role in the development of disease.

Key Words: Coronary artery disease, HSP27, networking biomarkers, risk prediction, Surface-Enhanced Laser Desorption/Ionization

Address for correspondence:
Dr. Rajani Kanth Vangala, Tata Proteomics and Coagulation Department, Thrombosis Research Institute, 258/A, Bommasandra Industrial Area, Anekal Taluk, Bangalore - 560 099, India. E-mail: rajani@triindia.org.in

Received: 13.07.2012, **Accepted:** 09.10.2012

Access this article online	
Quick Response Code:	Website: www.advbiores.net
	DOI: 10.4103/2277-9175.115805

INTRODUCTION

Coronary artery disease (CAD) is the principal cause of death in most countries and despite of major advances in treatment, a large number of victims die apparently healthy and suddenly without prior symptoms. The major challenge in cardiovascular medicine is to find a way of predicting the risk that

How to cite this article: Vangala RK, Ravindran V, Kamath K, Rao VS, Sridhara H. Novel network biomarkers profile based coronary artery disease risk stratification in Asian Indians. Adv Biomed Res 2013;2:59.

an individual will suffer from the disease.[1] Most risk prediction algorithms screen using the Framingham risk score (FRS), which considers conventional risk factors such as total cholesterol, high-density lipoprotein (HDL), smoking, hypertension, age, and gender in the algorithm. However, Kanjilal *et al*.[2] have shown that in Asian Indians, the Framingham model defined only 5% of their study cohort to be at high risk, which was an underestimation of CAD risk in the genetically predisposed population. Addition of new biomarkers may add a better value proposition to the risk prediction.[3] Furthermore, it was also shown that use of inflammatory markers such as C-reactive protein.[4] And coagulation factor-VII expression and genetic markers can add value for risk prediction in Asian Indians.[5] All data so far suggest that there is a need to identify better biomarkers to develop a comprehensive model for CAD risk prediction, especially for Indians.

Global proteome analysis can provide an overall understanding of the disease changes and contribute to the field of clinical cardiovascular science.[6,7] Biomarker discovery using Surface-Enhanced Laser Desorption/Ionization Mass Spectrometry (SELDI-TOF MS) is a novel approach and widely used in biomarker detection and identification.[8] This method is highly advantageous due to the sensitivity of the assay and the low sample volume requirement. Recent advances in the use of SELDI-TOF MS in CAD have been highlited by Wang,[6] in the chinese population.

In the present study we used the SELDI-TOF MS technology for identifying differentially expressed protein patterns in subjects with and without CAD. Furthermore, we have used three different techniques namely Support Vector Machine (SVM), Discriminant Analysis (DA), and Multilayer perceptron Artificial Neural Networks (ANN) for risk prediction. We identified that the SVM models can give better classification and therefore can be used along with protein profiles in risk prediction. Of the 9 *m/z* peaks, which could significantly discriminate affected and unaffected subjects, one of the peaks was HSP27 and was validated as a potential risk prediction biomarker in this study.

There are approximately 30,000 articles on cardiac biomarkers on PubMed. However, only a small number of these studies have yielded useful biomarkers for clinical purposes. Genes or proteins usually work collaboratively and involve several pathways. Protein-protein interactions and sub-networks play a major role in modulation of specific pathways and by using this information the predictive value of algorithms could be improved to higher levels. Based on the network profile developed from the biomarkers, we identified interaction of several pathways like stress (HSP27, DAOA), metabolic stress (ROMO1, QRFP), inflammation (INFA2, PLDN, CDKN2B, APP, FAU, and ENSG00000235915), coagulation (PLG, FGA, C3), obesity (APOC2, INSL4), hypertension (VIP), calcium binding (CALML4), and cell adhesion (VTN, MPZL3) as interacting members in the disease. The modulation of one or more of these pathways can lead to a chain reaction of changes in the pathways leading to the onset of CAD. Therefore, use of these novel biomarkers may give better risk prediction for CAD in Indians.

MATERIALS AND METHODS

Study participants and samples

The study comprised of 252 population based subjects out of which 112 probands without family history of CAD and 140 true controls were included. The baseline characteristics of study participants are shown in Table 1. The affected subjects were selected based on the following criteria: (1) Patient is a male ≤60 and female ≤65 on the onset of CAD, diagnosis of CAD via ECG/echo/biochemical or angiogram, patients posted for Percutaneous Transluminal Coronary Angioplasty (PTCA) and Coronary Artery Bypass Surgery (CABG) as diagnosed and given in the physicians report and also as answered in the questionnaire. The control subjects were enrolled above the age of 18 and should not have cardiovascular disease and other major illness like caner, liver failure according to the World Health Organization (WHO) guidelines. All the patient samples were collected after required ethics review board assessment and individual consent.

Biochemical assays

Blood was collected from the participants after a 12-h fasting period. Serum cholesterol and triglycerides were

Table 1: Baseline characteristics of study participants

Variables	Control group	CAD group	*P* value
n	112	140	
Age (years)	48.66±0.68	55.44±0.77	*P*<0.001
Gender (%)			
Male	78.6	92.0	0.003
Female	21.4	8.0	
BMI (kg/m^2)	24.70±0.34	25.02±0.38	0.05
Cigarette smoking (%)	43.7	56.3	0.02
Hypertension (%)	28.3	77.1	*P*<0.001
Diabetes (%)	8.5	91.5	*P*<0.001
Total cholesterol (mg/dl)	182.58±3.35	151.57±4.03	*P*<0.001
Triglycerides (mg/dl)	156.61±7.27	170.91±8.76	0.681
HDL (mg/dl)	43.45±0.81	37.50±0.98	*P*<0.001
LDL (mg/dl)	107.82±2.73	79.90±3.28	*P*<0.001

CAD: Coronary artery disease, Data are Mean±SE

estimated by standard enzymatic analyze following manufacturer's guidelines (Randox Laboratories, London, UK). HDL cholesterol was estimated after precipitation of non-HDL fractions with a mixture of 2.4 mmol/l phosphotungstic acid and 39 mmol/l magnesium chloride, and LDL cholesterol was estimated using the Friedewald formula[9]. A normal human serum pool (NHP) prepared in-house was run with each batch. The inter-assay coefficients of variation (CVs) for commercial controls and NHP ranged from 4.9% to 7% for total cholesterol, 6.1% to 7.7% for triglycerides, and 7.1% to 12.2% for HDL cholesterol.

Reagents and instruments

Sinapinic acid (SPA) and CM10 chip were purchased from Bio-Rad, Hercules, CA (USA) and all other reagents from Sigma Aldrich, St. Louis, MO (USA). The serum samples (in duplicates) were analyzed using CM10 chip followed by the Ciphergen Express Client software. Serum samples were thawed on ice and centrifuged at 14,000 r.p.m. for 5 min at 4°C. A 5-µl volume of supernatant of each sample and 10 µl of U9 buffer (9 M urea, 2% CAHPS, 1% dithiothreitol (DTT)) were added into a tube, which was mixed for 30 min on a platform shaker at 4°C. Next, 185 µl of sodium acetate (100 mM, pH 4) was added to the U9/serum mixture and mixed at 4°C for 2 min on the shaker. A 200-µl volume of sodium acetate was added and mixed for 5 min to activate the CM10 chips. Diluted samples (100 µl) were spotted onto a bioprocessor (Ciphergen Biosystems, Fremont, CA, USA) containing the ProteinChip arrays and then mixed on a platform shaker for 60 min at 4°C. The excess serum was discarded and the chips were washed three times with 200 µl of sodium acetate and twice with 5 µl of dH_2O. The chips were then removed from the bioprocessor and air-dried. Before SELDI analysis, 1 µl of a saturated solution of SPA (Bio-Rad) was applied on to each chip twice and air-dried.

Protein chip array analysis

A set of different protocols was used on each spot with varied laser intensities. Pre-processing was done using ProteinChip© software 3.1. The peaks with less background noise were considered for further analysis after baseline subtraction. After normalization peaks with standard deviation of ±2 were deleted. Finally clusters were made by Expression Differential Matrix (EDM) within a range of 1500-30,000 Da. We considered *m/z* less than 1500 as matrix noise. A first pass of 20% anda 0.3% mass window and a second pass of 2 were given. Mass accuracy was calibrated to <0.1% by all-in-one peptide molecular mass standard (Ciphergen Biosystems, Fremont, CA, USA).

Bioinformatics and biostatistics analysis

The baseline for the study participants were carried out using SPSS version 17 software. The continuous data were analyzed and calculated by Student's *t*-test and cross-tabs for the categorical data. We clustered the spectra and considered those spectra, which had a significant *P* value ($P < 0.05$) for further analysis. To better discriminate the CAD and control subjects based on the peak intensities for diagnostic profiling, we considered three methods, Support Vector Machine (SVM), Multilayer perceptron Artificial Neural Networks (ANN), and Discriminant Analysis (DA).[10,11] The type of SVM model we used was C-SVM and the kernel function used was RBF (radial basis function). Optimal values for parameters were found by SVM grid and pattern search with search criterion to minimize the total error. Each combination of peak was analyzed by 10-fold cross-validation. For ANN, architecture was made with an input layer with 36 neurons, a single hidden layer with nine neurons, and output layer with two neurons and four-fold cross-validation.

The 36 *m/z* peaks were determined as 31 potential biomarkers using proteomics tools from SWISSPROT (www.expasy.org) based on the mass and pI (standard deviation of ± 1% of the overall mass of the protein).[12-15]. We selected one peak (*m/z* 22859) corresponding to HSP27 for performing enzyme-linked immunosorbent assay (ELISA) assays in new set of affected ($n = 125$) and unaffected ($n = 431$) subjects. HSP27 ELISA (R and D Systems, Minneapolis, MN, USA; cat. no. DYC1580-2) was performed in serum samples of the subjects. The biomarkers identified above were given as input into STRING database (http://string-db.org/)[16] to generate the network of biomarkers for assessing functional association.

RESULTS

Out of 252 subjects, 91.5% of CAD-affected subjects were diabetic and 77.1% were suffering from hypertension [Table 1]. Furthermore the conventional risk factors hypertension, diabetes, smoking, total cholesterol, HDL, and age were found to be significant between cases and controls.

Differential protein pattern in controls and CAD-affected

The spectra of 112 subjects with CAD and 140 controls were analyzed. Fifty-six CAD samples and 70 control samples were used as test set and same number of samples in the as validation set for blind test. A total of 67 *m/z* clusters were obtained of which 36 were significantly ($P < 0.05$) differentially expressed. The specific proteins for each *m/z* were listed in Supplementary Table 1 after

the SWISSPROT database search. We obtained nine peaks that could discriminate the cases and controls in the SVM model. The descriptive statistics of these nine peaks are shown in Table 2. Biomarkers with *m/z* 22,859 [Figure 1], 9284, 14,660, 9481, and 14,720 were highly expressed in CAD-affected subjects, and *m/z* 5896, 8922, 8600, and 19,251 were highly expressed in controls [Table 2].

Comparison of three different approaches of model building

The 36 peaks were further analyzed by different techniques to obtain the best set of peaks and algorithm for risk prediction. We compared performance of three algorithms Discriminative Analysis (DA), Multilayer perceptron Artificial Neural Networks (ANN), and Support Vector Machine (SVM) based on accuracy, sensitivity, specificity, and area under the receiver operating curve (ROC) [Table 3a]. SVM was found to be the best model for classification using our data with an area under the curve (AUC) of 0.807 and better specificity, sensitivity, and accuracy. Furthermore, the test set also gave good classification data with SVM [Table 3b] with AUC of 0.785 and other features. Also when we consider the overall misclassification the least values were observed for SVM with 23.02% and 26.19% for training and test data, respectively.

Use of SELDI biomarkers and modulation of seven different pathways for risk stratification

As we know that FRS is used widely for risk prediction; however we also know that the use of FRS is limited for Asian Indians. Therefore, we considered the nine peaks identified by SELDI-TOF-MS as potential biomarkers along with the FRS model for risk stratification. FRS alone gave an AUC of 0.888, which improved to 0.963 on addition of the nine potential biomarkers [Figure 2]. These nine biomarkers represent seven different pathways, stress and stress/immunity (*m/z* s 22,859: HSP27, 5896: Leukocyte-specific transcript-1), coagulation (*m/z* s 8922: Plasminogen precursor activating peptide, 9284: Vitronectin-10, 8600: Pallidin gene isoform-2), infection and inflammation

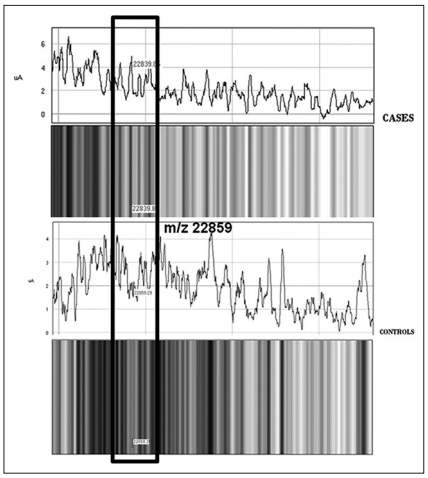

Figure 1: Representative spectrum report of average *m/z* 22,859 in CAD and control samples

Table 2: Mean intensity±SE levels in CAD and controls for biomarkers in the test data

Peaks	Identified protein	CAD-affected	Controls	*P* value
22,859	HSP27	4.19±0.14	3.41±0.13	*P*<0.001
5896	Leukocyte-specific transcript-1 protein (LST1_HUMAN)	122.63±8.37	196.33±7.83	0.023
8922	Plasminogen precursor-activating peptide (PLMN_HUMAN)	323.27±14.54	367.84±13.61	0.026
9284	Vitronectin V10 subunit (VTNC_HUMAN)	504.68±18.50	391.53±17.32	*P*<0.001
19,251	Interferon α2 Chain (IFNA2_HUMAN)	2.36±0.37	4.49±0.35	*P*<0.001
14,660	Farataxin chain 3 Farataxin (78 - 210) (FRDA_HUMAN)	16.71±0.57	14.56±0.53	0.006
9481	Calmodulin-like protein-4 isoform-3 (CALL4_HUMAN)	94.72±2.74	82.02±2.57	0.001
14,720	Cyclin-dependant kinase-4 inhibitor-B (CDN2B_HUMAN)	7.60±0.22	6.92±0.20	0.021
8600	Pallidin gene isoform-2 (PLDN_HUMAN)	120.63±8.38	151.80±7.50	0.006

CAD: Coronary artery disease

Table 3a: Classification of CAD and controls using three different methods, SVM, ANN, and DA, suggesting that SVM model is the best classifier for training data

Model	Accuracy (%)	Sensitivity (%)	Specificity(%)	AUC
SVM				
Training	76.98	82.86	69.64	0.807
ANN				
Training	76.98	82.86	69.64	0.794
DA				
Training	73.81	77.14	69.64	0.776

ANN: Multilayer perceptron Artificial Neural Networks, AUC: Area under the curve, CAD: Coronary artery disease, DA: Discriminant analysis, SVM: Support Vector Machine

Table 3b: Classification of CAD and controls using three different methods, SVM, ANN, and DA, suggesting that the SVM model is the best classifier for training data

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
SVM				
Validation	73.81	85.71	58.93	0.785
ANN				
Validation	68.25	78.57	55.36	0.765
DA				
Validation	70.63	80.00	58.93	0.741

ANN: Multilayer perceptron Artificial Neural Networks, AUC: Area under the curve, CAD: Coronary artery disease, DA: Discriminant analysis, SVM: Support vector machine

(*m/z* 19,251: Interferon α-2), mitochondrial damage (*m/z* 14,660: Farataxin chain-3), calcium binding (*m/z* 9481: Calmodulin-like protein-4 isoform-3), and cell cycle (*m/z* 14,720: Cyclin-dependent kinase-4 inhibitor-B). Modulation of these biomarkers results in the change in the functional implications of the pathways, which may result in the disease.

Networking biomarkers and pathways

The proteins identified [Supplementary Table 1] were further taken to generate the functional association network among themselves and with the other proteins. These proteins are from multiple pathways like inflammation, cell signaling, cell adhesion, immunity, obesity, lipid metabolism, coagulation, stress, membrane transport, protein degradation, coagulation, and cell cycle. Our data suggest that CAD is a multi-factorial process and deregulation of these factors may lead to the disease. As seen in Figure 3a, the network of the proteins identified suggests that 16 of 31 proteins have minimal or no linkage among themselves (MPZL3, INSL4, SCOC, ROMO1, CALML4, ENSG0000023591, PSAP, SRP9L1, ANKDD1A, CMTM1, CDN2B, HMSDV, DSC10, KRTDAP, VGLL4, FXN, and APOC2). These proteins need to be proven further to understand their biological role suggesting that they might be novel biomarkers for CAD in this study. Fifteen proteins were networked with at least one more protein (HSPB1, FGA, PLG, VTN, APP, QRFP, C3, POMC, CHGA, VIP, CRH, FAU, C11orf10, IFNA2, and CDKN2B).

Further, when we extended this network by adding other interacting or functionally associated proteins [Figure 3b], we saw that the individual proteins, which were not in the network in Figure 3a, had changed from 17 to 9 proteins. These nine proteins (ANKDD1A, ROMO1, SCOC, SRP9L, MPZL3, CALML4, CMTM1, KRTDAP, and ENSG0000023591) are potentially novel members and further analysis may be needed to identify their networks and associations. However, most of the other biomarkers were directly or indirectly were associated.

CAD-associated networks

Our data [Figure 3a and b] suggest that multiple pathways are associated and networked together in the CAD subjects for the onset of the disease. The most networked proteins [Supplementary Table 1] were identified based on number of edges for each protein in the network [Figure 3b]. The biomarkers FAU, CRH, APP, VIP, CHGA, POMC, HSPB1, C3, FGA, VTN, INFA2, FXN, CDKN2B, and PLDN are from different pathways suggesting that these pathways interact in the disease condition. These association studies suggest that coagulation, cell signaling, kinase inhibitors, stress, protease inhibitors, and obesity are major pathways leading to CAD in Asian Indians

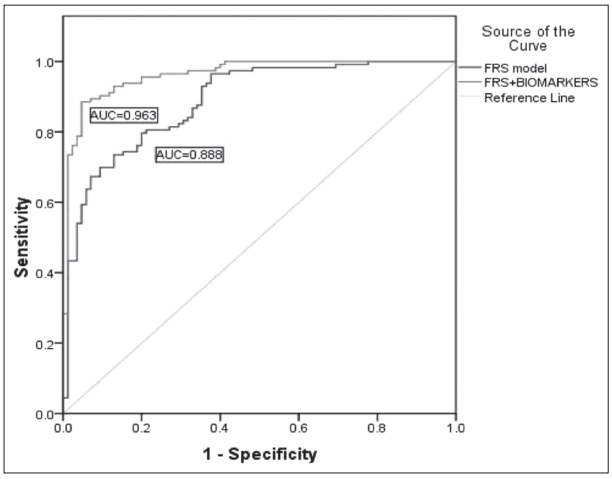

Figure 2: Receiver operating curves for FRS and for addition of biomarker expressions in discriminating CAD vs. controls. The improvement of AUC curve suggests that addition of SELDI-TOF-based feature selection biomarkers may add value in CAD risk stratification

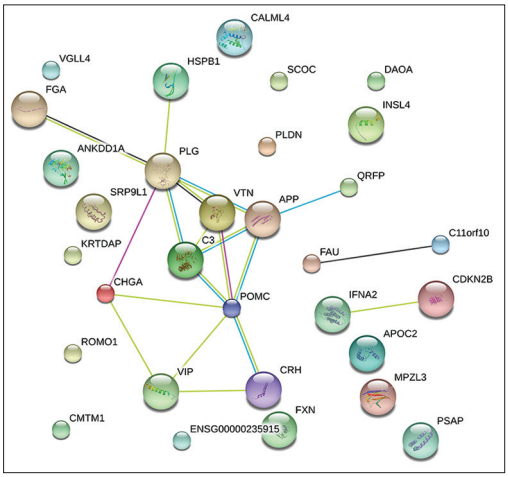

Figure 3a: Functional association of proteins identified by SELDI-TOF MS

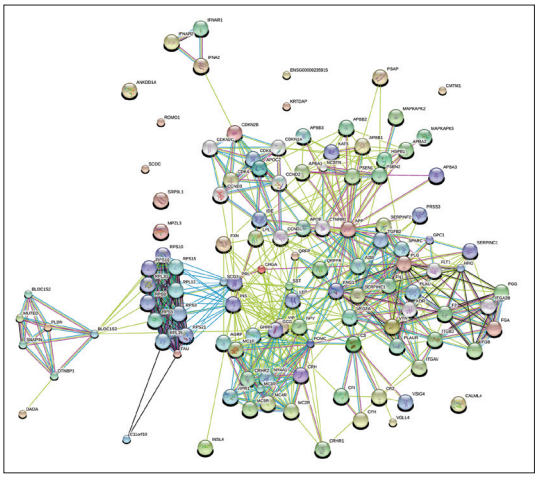

Figure 3b: Extended network and association of potential biomarkers

Supplementary Table 1: m/z peaks identified after SELDI-TOF analysis. Significant peaks used for discrimination of CAD and controls are marker in bold.

S. NO	Average M/Z	Protein Name	Pathway	Functional Associations
1	4467.4	Transcription cofactor vestigial-like protein 4 VGLL4_HUMAN (VGLL4)	Transcription Factor Co-activator Vestigial like 4 (Drosophila); May act as a specific coactivator for the mammalian TEFs (By similarity) (296 aa)	VEGFA
2	22676	Myelin protein zero-like protein 3 MPZL3_HUMAN (MPZL3)	Cell Adhesion Myelin protein zero-like 3; Mediates homophilic cell-cell adhesion (By similarity) (235 aa)	—
3	9430.8	Keratinocyte differentiation-associated protein KTDAP_HUMAN	Differentiation	—
4	7756.9	Ubiquitin-like protein FUBI UBIM_ HUMAN (FAU)	Ubiquitin like function	RPS3, RPL35, RPS5, RPS10, RPS21, C11ORF10, RPS16, RPL30, RPS9, RPS15, RPL12, INS, PRL
5	5896.5	Leukocyte-specific transcript 1 protein LST1_HUMAN (ENSG0000023591)	Cell shape and Immunity	—
6	9080.8	UPF0197 transmembrane protein C11orf10 CK010_HUMAN (C11orf10)	Membrane Transport	FAU, RPL35S
7	2789.8	Early placenta insulin-like peptide INSL4_HUMAN (INSL4)	Cell Signaling Insulin-like 4 (placenta); May play an important role in trophoblast development and in the regulation of bone formation (139 aa)	INS
8	2952.2	CKLF-like MARVEL transmembrane domain-containing protein 1 CKLF1_ HUMAN (CMTM1)	Membrane Transport	—
9	4643.5	Ankyrin repeat and death domain-containing protein 1A, Isoform 3, AKD1A_HUMAN (ANKDD1A)	Signal Transduction	—
10	4746.2	Corticoliberin CRF_HUMAN (CRH)	Inflammatory respnse	PRL, GHRH, NPY, MC2R, LEP, SST, AGRP, MC4R, CRHR1, MC5R, VIPR1, CRHR2, POMC, NPY, MC1R, GCG, VEGFA, NR4A1, MC3R, VIP
11	5918.6	Gamma-secretase C-terminal fragment 50, A4_HUMAN (APP)	Protease Inhibitor	APBA1, APBB1, APOE, APBB3, QRFPR, KAT5, PSEN1, A2M, PSEN2, APBB2, GCG, SPARC, SST, GPC1, POMC, KNG1, NPY, PLG, NCSTN, PRS53, C3, APBA2, IDE, FN1, APBA3, SERPINF2, SERPINE1, QRFP, TGFB2
12	6088.6	Pro-opiomelanocortin, (Lipotropin gamma) COLI_HUMAN (POMC)	Obesity, cell signalling	PRL, MC5R, FNI, MC1R, MC3R, GHRH, AGRP, VTN, MC2R, NPY, MC4R, APP, GCG, VIPR1, CHGA, VIP, NR4A1, SCG3, KNG1, CHR2, C3, A2M, SST, CRH, LEP, CRHR1, VEGFA, INS
13	6108.8	Short coiled-coil protein Isoform 2, SCOC_HUMAN	Protein Binding, Cell signaling	—
14	9137.3	Proactivator polypeptide (Saposin-D), SAP_HUMAN (PSAP)	Lipid Metabolism and Glycosilation	PSEN2, PSEN1
15	9287.4	Down syndrome critical region protein 10, DSC10_HUMAN	Uncharacterized	—
16	4538.2	VIP peptides, Intestinal peptide PHV-42, VIP_HUMAN (VIP)	Vasodilation, lowers arterial blood pressure, stimulates myocardial contractility, increases glycogenolysis and relaxes the muscles	CRHR2, GCG, LEP, CCND1, PRL, SST, POMC, VPR1, CHGA, SCG3, KNG1, VEGFA, NPY, INS, CRH
17	4522	Orexigenic neuropeptide QRFP, QRF-amide, OX26_HUMAN (QRFP)	Stimulates feeding behavior, metabolic rate and locomotor activity and increases blood pressure. May have orexigenic activity. May promote aldosterone secretion by the adrenal gland	LEP, KNG1, GCG, NPY, QRFPR, APP
18	4636.7	Ankyrin repeat and death domain-containing protein 1A, Isoform 3, AKD1A_HUMAN (ANKDD1A)	Signal Transduction	0
19	5894.6	Reactive oxygen species modulator 1, Isoform 2, ROMO1_HUMAN (ROMO1)	Oxidative stress	0

Table 1 (contd...)

Table 1 (contd...)

S. NO	Average M/Z	Protein Name	Pathway	Functional Associations
20	6031.3	Minor histocompatibility protein HMSD variant form, HMSDV_HUMAN	Precursor of the histocompatibility antigen ACC-6	0
21	8918.5	Apolipoprotein C-II, APOC2_HUMAN (APOC2)	Lipid binding	LPL. APOBB3, APOE, APOC2
22	9124	Isoform 2 of Signal recognition particle 9 kDa protein, SRP09_HUMAN (SRP09L1)	Signal-recognition-particle assembly has a crucial role in targeting secretory proteins to the rough endoplasmic reticulum membrane. SRP9 together with SRP14 and the Alu portion of the SRP RNA, constitutes the elongation arrest domain of SRP. The complex of SRP9 and SRP14 is required for SRP RNA binding.	0
23	1894.1	Isoform 2 of D-amino acid oxidase activator, DAOA_HUMAN	activate D-amino acid oxidase	DTNBP1
24	2785.7	Early placenta insulin-like peptide A chain, INSL4_HUMAN (NSL4)	Cell Signalling	INS
25	5086.6	Pancreastatin, Chromogranin-A, CMGA_HUMAN (CHGA)	strongly inhibits glucose induced insulin release from the pancreas	NPY, GCG, SST, SCG3, PLG, POMC, VIP, INS
26	6083.7	Lipotropin gamma, Pro-opiomelanocortin, COLI_HUMAN (POMC)	Hormonal balance in skin pigmentation	PRL, MC5R, FN1, MC1R, MC3R, GHRH, AGRP, VTN, MC2R, NPY, MC4R, APP, GCG, VIPR1, CHGA, VIP, NR4A1, SCG3, KNG1, CRHR2, C3, A2M, SST, CRH, LEP, CRHR1, VEGFA, INS
27	22859	HSP27 HSPB1_HUMAN (HSPB1)	Stress and immune response	CCND1, PLG, CDKN1A, CTNND1, MAPKAPK5, MAPKAPK2
28	8922	Plasminogen Precursor Activating Peptide PLMN_HUMAN (PLG)	Involved in coagulation	SERPINF2, SERPINE1, FN1, C3, KNG1, SPARC, HSPB1, VTN, TGFB2, APP, FGG, FGB, FGA, A2M, PLAU, PLAUR, HRG, SERPINC1, IGGB3
29	8900	Compliment C3a CO3_HUMAN (C3)	Involved in coagulation	PLG, APOE, CR2, INS, VTN, VSIG4, CFI, CFH, APP. POMC, KNG1, FNI, NPY,SST
30	5901.57	Fibrinogen alpha-E chain decomposition product FIBA_HUMAN (FGA)	Involved in coagulation. Fibrinogen has a double function: yielding monomers that polymerize into fibrin and acting as a cofactor in platelet aggregation	SERPINF2, FGB. SERPINE1, ITGA2B, PLG, ITGB3, F2, HRG, FN1, FGG
31	9284	Vitronectin V10 subunit VTNC_HUMAN (VTN)	Interacts with PAI1 in coagulation pathway. Vitronectin is a cell adhesion and spreading factor found in serum and tissues. Vitronectin interact with glycosaminoglycans and proteoglycans. Is recognized by certain members of the integrin family and serves as a cell-to-substrate adhesion molecule. Inhibitor of the membrane-damaging effect of the terminal cytolytic complement pathway.	SERPINC1, KNG1, ITGAV, POMC, SERPINE1, ITGA2B, FGG, PLG, FNI, KDR, C3, PLAU, HRG, TGFB2, VEGFA, ITGB3, SPARC
32	19251	Interferon Alpha 2 Chain IFNA2_HUMAN (INFA2)	Inflammation. Produced by macrophages, IFN-alpha has antiviral activities.	CDKN2B, IFNAR1, INFAR2
33	14660	Farataxin Chain 3 Frataxin(78-210) FRDA_HUMAN (FXN)	Mitochondrial damage	LEP, INS, LPL, NPY, APBA1, AGRP, GCG
34	9481	Calmodulin like protein 4 Isoform 3 CALL4_HUMAN (CALML4)	Calcium ion binding	–
35	14720	Cyclin dependant kinase 4 inhibitor B CDN2B_HUMAN	Cell cycle. Interacts strongly with CDK4 and CDK6. Potent inhibitor. Potential effector of TGF-beta induced cell cycle arrest.	IDE, CTNNB1, CDK4, CDKN1A, CDKN2B, CCND1, CCND3, IFNA2, CDKN2C, CDK6
36	8600	Pallidin Gene Isoform 2 PLDN_HUMAN (PLDN)	Platelet storage pool deficiency. Involved in the development of lysosome-related organelles, such as melanosomes and platelet-dense granules. May play a role in intracellular vesicle trafficking, particularly in the vesicle-docking and fusion process.	BLOC1S2, MUTED, SNAPIN, DTNBP1, BLOC1S3

in our studies. It is understood that stress leads to several changes, which might play a major role in early pathogenesis of CAD. Therefore, markers like ROMO1 and HSP27 might be the first candidate markers, which need to be evaluated, and we evaluated HSP27 as potential marker for the same.

The stress-related protein HSP27 is highly associated with CAD in Asian Indians

HSP27 is a member of the small heat-shock protein (HSP) (sHSP) family and is involved in diverse range of functions in addition to its chaperoning function. We identified *m/z* 22,859 as HSP27 (molecular weight 22,783 Da). HSP27 is also a member associated with five different partners, which in turn are regulating multiple pathways such as PLG (plasminogen precursor-activating peptide), CDKN1A, CTNNB1, CCND1 (cell-cycle proteins), and kinases such as MAPKAPK2 and 5. These associations suggest that many cellular responses may be triggered along with HSP27 in CAD and therefore this biomarker was further validated. We performed ELISA assay for HSP27 in 431 subjects (125 CAD-affected and 306 unaffected). We found that the CAD-affected subjects had higher expression levels that unaffected [Table 4a]. The odds ratio of HSP27 alone [Table 4b] was not significant; however, after addition of conventional risk factors (age, gender, body mass index, waist circumference, and hypertension), the odds ratio of fourth quartile in comparison to first quartile improved to 2.81 (95% confidence interval (CI): 1.18-6.79, $P = 0.019$). Furthermore, upon adjustment with lipids (triglycerides, total cholesterol, HDL, and LDL), the odds ratio of the fourth quartile improved to 3.47 (95% CI: 1.41-8.56, $P = 0.007$).

DISCUSSION

As CAD is a major killer in India, it is very important to identify the ways of improving risk prediction. At present diagnosis or risk prediction is dependent on clinical history, physical examination, and other tests, which do not look at the biochemical or molecular changes, which might give early risk prediction to CAD. In our present study, we have explored and validated the process of identification of biomarkers using SELDI-TOF-MS and further using the patterns of *m/z* to diagnose the risk of CAD in Asian Indians. It has been well established that SELDI-TOF-based CAD diagnosis can be used;[6-10] our attempt to use novel biomarkers for risk prediction can add value for early diagnosis and prevention of CAD endemic.

SELDI-TOF-based biomarker detection and use of protein patterns derived from serum to differentiate between CAD and no CAD is of major interest[11,17] as it allows complete proteome profiling in a high-throughput format. Using feature selection techniques such as SVM helps in a more robust method of classifying the subjects.[18,19] Our analysis showed that the power of each biomarker to discriminate between the cases and controls was best for the SVM model by estimating the ROC. The greater the AUC value for the biomarkers shows the relative importance value of the ability to accurately distinguish between the different groups.[20] Apart from using the SVM to identify the best biomarkers and their discrimination between CAD and no CAD, we have also established that the biomarkers add a very important value for FRS risk prediction method.

The 31 proteins identified and their networking suggests that multiple pathways are associated with CAD and in specific inflammation, cell signaling, coagulation, cell adhesion, stress, and obesity are the major pathways. It is also very interesting to note novel proteins, which are not associated with any of the known pathways to be identified [Figure 3a and b]. These proteins may have no earlier data in relation to coronary artery disease and more studies may be needed to understand their role. Of the 15 proteins identified as highly networked FAU, CRH, POMC, VIP, VTN, and PLG are prominent with more than 10 associations suggesting that further analysis with these proteins may yield better understanding of the pathways involved in the CAD.

Table 4a: Mean expression levels of HSP27 in CAD-affected and unaffected subjects

Biomarker	Affected	Unaffected	*P* value
HSP27	1376.95±210.41	854.48±35.08	0.016

CAD: Coronary artery disease, Mean±SE

Table 4b: Association of HSP27 based on odds ratio

	HSP27 quartiles	1	2	3	4
Model 0	Odds ratio (95% CI)	1	1.39 (0.78-2.48)	0.66 (0.35-1.23)	1.23 (0.69-2.21)
Quartiles alone	*P* value	0.091	0.261	0.191	0.49
Model 1	Odds ratio (95% CI)	1	2.59 (1.10-6.09)	1.48 (0.58-3.79)	2.84 (1.18-6.79)
Adjusted with CRF	*P* value	0.066	0.029	0.416	0.019
Model 2	Odds ratio (95% CI)	1	2.48 (1.03-5.97)	1.41 (0.53-3.74)	3.47 (1.41-8.56)
Additional adjustment with lipids	*P* value	0.034	0.043	0.485	0.007

HSP27 shows good association when the model is adjusted for conventional risk factors (body mass index, hypertension, waist circumference, age, and gender) and lipids (triglycerides, total cholesterol, HDL, and LDL)

Our data suggest that the stress- and immunity-related protein HSP27 might play a major role in CAD for Asian Indians [Table 4a and b]. The functional role of several HSPs in atherosclerosis has been well studied as they represent the response of the cells to blood vessel to different stress signals.[21] It is also well known that HSPs are potential targets for immune response and contribute to inflammatory process.[22] The smooth muscle cells (SMCs) play a important role in atherogenesis as they can over express HSPs as a part of survival mechanism following exposure to variety of stressors (example: High blood pressure). Most research on HSPs was focused on HSP60/65 and 70; however recently evidence of role of HSP27 in CAD is becoming evident.[23-26] In our study we have analyzed the serum levels of 431 subjects and found that HSP27 alone is not associated, but when the model was adjusted for conventional risk factors and lipids, higher association to CAD was seen [Table 4b]. Our data suggest that HSP27 might play important role in risk prediction and further studies are needed to evaluate the value addition by this biomarker. It was also suggested that phosphorylated HSP27 may have a protective effect in atherogenesis;[27] therefore further studies are needed to evaluate the functional role of HSP27 versus the use of expression levels in risk prediction.

Our findings revealed that SELDI-TOF-MS technique can be used for risk stratification of CAD-affected and unaffected subjects using the SVM method. The networking of proteins and the pathways indicate that several pathways such as stress, inflammation, coagulation, cell adhesion, signaling, and obesity are interlinked and might crosstalk in the development of the disease. Our approach has resulted in understanding the network and modulation of pathways that contain specific sub-networks and novel biomarkers that may help in improving the risk prediction. Further, we used SELDI-TOF-MS not only in identification of new biomarkers, but also as a means of understanding the mechanism of CAD development by network construction.

ACKNOWLEDGMENT

This work was supported by Department of Biotechnology, Government of India (grant no. BT/01/COE/08/07). The authors also would like to thank the faculty and other scientists from Thrombosis Research Institute, Bangalore, India for help in the research.

REFERENCES

1. Martin-Ventura JL, Blanco-Colio LM, Tuñón J, Muñoz-García B, Madrigal-Matute J, Moreno JA, *et al.* Biomarkers in cardiovascular medicine. Rev Esp Cardiol 2009;62:677-88.
2. Kanjilal S, Rao VS, Mukherjee M, Natesha BK, Renuka KS, Sibi K, *et al.* Application of cardiovascular disease risk prediction models and the relevance of novel biomarkers to risk stratification in Asian Indians. Vasc Health Risk Manag 2008;4:199-211.
3. Kim HC, Greenland P, Rossouw JE, Manson JE, Cochrane BB, Lasser NL, *et al.* Multimarker prediction of coronary heart diasease risk: The Women's Health Initiative. J Am Coll Cardiol 2010;55:2080-91.
4. Rao VS, Kadarinarasimhiah NB, John S, Hebbagodi S, Shanker J, Kakkar VV. Usefulness of C-reactive protein as a marker for prediction of future coronary events in the Asian Indian population: Indian atherosclerosis research study. Int J Vasc Med 2010;2010:389235.
5. Shanker J, Perumal G, Maitra A, Rao VS, Natesha BK, John S, *et al.* Genotype-phenotype relationship of F7 R353Q polymorphism and plasma factor VII coagulant activity in Asian Indian families predisposed to coronary artery disease. J Genet 2009;88:291-7.
6. Wang Y, Hu X, Wang JA, Xiang M, Yu J, Liu F, *et al.* Application of serum protein fingerprint in diagonisis of coronary artery disease. Clin Biochem 2011;44:185-91.
7. Edwards AV, White MY, Cordwell SJ. The role of proteomics in clinical cardiovascular biomarker discovery. Moll Cell Proteomics 2008;7:1824-37.
8. Issaq HJ, Veenstra TD, Conrads TP, Felschow D. The SELDI-TOF MS approach to proteomics: Protein profiling and biomarker identification. Biochem Biophys Res Commun 2002;292:587-92.
9. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972;18:499-502.
10. Jin G, Zhou X, Wang H, Zhao H, Cui K, Zhang XS, *et al*. The knowledge-integrated network biomarkers discovery for major adverse cardiac events. J Proteome Res 2008;7:4013-21.
11. Chen YD, Zheng S, Yu JK, Hu X. Artificial neural networks analysis of surface enhanced laser desorption/ionization mass spectra of serum protein pattern distinguishes colorectal cancer from healthy population. Clin Cancer Res 2004;10:8380-5.
12. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al.* Protein identification and analysis tools on the ExPASY server. In: Walker JM, editor. The Proteomics Protocols Handbook.Humana Press Inc (Springer Protocols). New York, USA; 2005. p. 571-607.
13. Wilkins MR, Gasteiger E, Sanchez JC, Appel RD, Hochstrasser DF. Protein identification with sequence tags. Curr Biol 1996;6:1543-4.
14. Wilkins MR, Gasteiger E, Tonella L, Ou K, Tyler M, Sanchez JC, *et al.* Protein identification with N and C-terminal sequence tags in proteome projects. J Mol Biol 1998;278:599-608.
15. Lomthaisong K, Bonmaleerat K, Wingpia A. Proteomic study of recombinant *Escherichia coli* expressing *Beauveria bassiana* chitinase gene. Chiang Mai J Sci 2008;35:324-30.
16. Szklarczyk D, Franceschini A, Khun M, Simonovic M, Roth A, Minquez P, *et al.* The STRING database in 2011: Functional interaction networks of proteins globally integrated and scored. Nucleic Acid Res 2011;39:D561-8.
17. Baggerly KA, Morris HS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: Comparing datasets from different experiments. Bioinformatics 2004;20:777-85.
18. Willian SN. What is support vector machine? Nat Biotechnol 2006;24:1565-7.
19. Son YJ, Kim HG, Kim EH, Choi S, Lee SK. Applications of support vector machine for prediction of medication adherence in heart failure patients. Healthc Inform Res 2010;16:253-9.
20. Yu JK, Zheng S, Tang Y, Li L. An integrated approach utilizing proteomics and bioinformatics to detect ovarian cancer. J Zhejiang Univ Sci B 2005;6:227-31.
21. Wick G, Schett G, Amberger A, Kleindienst R, Xu Q. Is athreosclerosis an immunologically mediated disease? Immunol Today 1995;16:27-33.
22. Ghayour-Mobarhan M, Rahsepar AA, Tavaillaie S, Rahsepar S, Ferns GA. The potential role of heat shock proteins in cardiovascular disease: Evidence from *in vitro* and *in vivo* studies. Adv Clin Chem 2009;48:27-72.
23. Wick G. The geat is on: Heat shock proteins and atherosclerosis. Circulation 2006;114:870-2.

24. Galis ZS, Sukhova GK, Lark MW, Libby P. Increased expression of matrix metalloproteinases and matrix degrading activity in vulnerable regions of human atherosclerosis plaques. J Clin Invest 1994;94:2493-503.
25. Lijnen HR. Plasmin and matrix metalloproteinases in vascular remodelling. Thromb Haemost 2001;86:324-33.
26. Martin-Ventura JL, Duran MC, Blanco-Colio LM, Meilhac O, Leclercq A, Michel JB, *et al.* Identification by a differential proteomic approach of heat shock protein 27 as a potential marker of atherosclerosis. Circulation 2004;110:2216-9.
27. Hollander JM, Martin JL, Belke DD, Scott BT, Swanson E, Krishnamoorthy V, *et al*. Overexpression of wild-type heat shock protein 27 and a nonphosphprylatable heat shock protein 27 mutant protects against ischemia/reperfusion injury in a transgenic mouse model. Circulation 2004;110:3544-52.

Source of Support: This work was supported by Department of Biotechnology, Government of India (grant no. BT/01/COE/08/07)., **Conflict of Interest:** None declared.