

# Danger-recognizing proteins, $\beta$ -defensin-128 and histatin-3, as potential biomarkers of recurrent coronary events

MADANKUMAR GHATGE<sup>1,2\*</sup>, ANKIT SHARMA<sup>2,3</sup>, SANGEETA MAITY<sup>1</sup>,  
VIJAY V. KAKKAR<sup>4</sup> and RAJANI KANTH VANGALA<sup>1,3\*</sup>

<sup>1</sup>Tata Proteomics and Coagulation Unit, Thrombosis Research Institute, Bangalore 560099; <sup>2</sup>Manipal University, Madhav Nagar, Manipal 576104, Karnataka; <sup>3</sup>Elizabeth and Emmanuel Kaye Bioinformatics and Statistics Unit, Thrombosis Research Institute, Bangalore 560099, India; <sup>4</sup>Thrombosis Research Institute, London SW3 6LR, UK

Received February 17, 2016; Accepted May 11, 2017

DOI: 10.3892/ijmm.2017.3031

**Abstract.** Conventional risk factors have limited ability to predict recurrent events in subjects with first-time coronary artery disease (CAD). This aim of this study was to identify novel biomarkers using comparative global proteome analysis to improve the risk assessment for recurrent coronary events. We used samples from phase-I of the Indian Atherosclerosis Research Study (IARS), consisting of 2,332 subjects, of whom 772 were CAD-affected subjects, including 152 with recurrent events identified during a 5-year follow-up period. Global proteome analysis was performed on serum samples of 85 subjects with recurrent coronary events and 85 age- and gender-matched subjects with first-time CAD using surface-enhanced laser desorption ionization time-of-flight mass spectrometry with CM10 arrays. TagIdent was used for protein identification followed by validation by western blot analysis and ELISA. Data were analyzed by logistic analysis, Cox-regression, hazards ratio, C-statistics and combined-marker risk score using SPSS version-17 and R-package version-2.13.0 software. We identified 16 significantly differentially expressed protein peaks. Of these, 2 peaks corresponding to m/z 8588 and 1864 were identified as  $\beta$ -defensin-128 and histatin-3, belonging to the danger-recognizing peptide family, which exhibited a significant independent association with recurrent events (odds ratios of 7.49 and 1.4, respectively). C-statistics improved significantly from 0.677 for conventional risk factors alone to 0.800 (p-value=0.001) in combination with  $\beta$ -defensin-128 and histatin-3 with a hazards ratio of 1.833. A combined risk score of  $\beta$ -defensin-128 and histatin-3 could reclassify 112 out of the

170 subjects into intermediate- and high-risk groups. On the whole, our data indicate that  $\beta$ -defensin-128 and histatin-3 may be potential biomarkers which may be used to improve risk the stratification for recurrent coronary events.

## Introduction

Cardiovascular diseases have reached epidemic proportions in India, particularly among the younger population in comparison to other ethnic groups (1). Patients who have survived a first-time coronary event are at a greater risk of developing recurrent episodes, as compared to asymptomatic subjects (2). Several biomarkers belonging to different pathways have been identified in coronary artery disease (CAD) and are being deployed to identify the subset of subjects who are at a greater risk of developing acute coronary events. These biomarkers are associated with inflammation (C-reactive protein), hemodynamic stress (brain natriuretic peptide, N-terminal prohormone brain natriuretic peptide) and kidney dysfunction (albuminuria and cystatin-C) (3,4). Pathogen burden caused due to the simultaneous infection by multiple pathogens, such as *Chlamydia pneumoniae*, cytomegalovirus (CMV), *Helicobacter pylori*, Herpes simplex virus (5) and in particular, neutralizing antibody titers against CMV along with inflammatory markers have been shown to be important for risk stratification in the Asian Indian population (6).

On the other hand, the use of global proteome analysis can not only provide a better understanding of the underlying disease process, but can also help to unearth novel disease markers that can enhance the power to predict the risk of acute coronary events when combined with other well-studied markers as discussed above (7). This includes the comparative proteomics approaches based on mass spectrometric analysis of protein profiles in serum samples obtained from normal and diseased subjects. Therefore, the aim of this study was to perform global proteome analysis using the surface-enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI-TOF MS) platform to possibly identify novel biomarkers associated with recurrent coronary events in patients with established CAD. We identified 2 infection-related proteins, namely  $\beta$ -defensin-128 and histatin-3 as potential biomarkers for recurrent CAD.

---

*Correspondence to:* Dr Rajani Kanth Vangala, Tata Proteomics and Coagulation Unit, Thrombosis Research Institute, 258/A, Bommasandra Industrial Area, Anekal Taluk, Bangalore 560099, India  
E-mail: rajani@triindia.org.in

\*Contributed equally

**Key words:** recurrent coronary artery events, global proteome analysis,  $\beta$ -defensin, histatin-3, risk prediction

## Materials and methods

**Study population.** The Indian Atherosclerosis Research Study (IARS) is a prospective cohort study designed to investigate and understand the molecular basis of CAD in the Indian population. The design and details of data collection and the procedures used are described elsewhere (8). The study was designed in accordance with the principles and the guidelines of World Medical Association Declaration of Helsinki and the Indian Council of Medical Research and was approved by the Thrombosis Research Institute Ethics Committee. The participants provided their signed informed consent to participate in the study. In phase-I of this study between 2004-2006, 2,332 subjects were enrolled of which 772 (33%) were CAD-affected. A total of 152 subjects from this CAD-affected group experienced a recurrence of cardiac events during the course of a 5-year follow-up period. All the subjects in this study had a family history of CAD, and we matched the age and gender of 85 subjects in the recurrent events group to those of 85 subjects with first-time events, and the serum samples of these subjects was used for global proteome analysis. The baseline characteristics of the study participants are presented in Table I. The subjects were recruited based on the following criteria: i) age at onset of CAD was  $\leq 60$  years for males and  $\leq 65$  years for females; ii) the diagnosis of CAD was based on an electrocardiogram (ECG)/echocardiography (echo)/biochemical or angiogram findings; iii) patients who were posted for percutaneous transluminal coronary angiography (PTCA)/coronary artery bypass graft (CABG). The subjects who did not meet the above-mentioned criteria or had any other major illness, as defined by the World Health Organization, such as cancer, liver failure, etc. were excluded from the study. All blood samples were collected after overnight fasting for 12-14 h, and the plasma and serum aliquots were stored at  $-80^{\circ}\text{C}$ . Detailed information on medical history, demographics and relevant information, including 3-generation pedigree was recorded in a questionnaire after a personal interview. The diagnosis of diabetes and hypertension was based on medical reports, while the body mass index was calculated.

**Telephonic assessment of cardiac health status.** The periodic telephonic follow-up of all study participants was undertaken to obtain an update on the cardiac health status and 5 such rounds of follow-up were completed. Follow-up was performed once every 18 months on an average from the time of recruitment with a successful follow-up rate of 60%. Endpoints included non-fatal events (re-do PTCA/re-do CABG/heart attack, stroke) and fatal events (fatal heart attack/fatal stroke/sudden death/cardiac arrest). This information was verified through either paper trail or hospital records whenever available.

**Global proteome analysis.** Global proteome analysis of serum samples was performed using SELDI-TOF MS. Sinapinic Acid (SPA) and CM10 chip were purchased from Bio-Rad (Hercules, CA, USA) and all the other reagents were from Sigma-Aldrich (St. Louis, MO, USA). Serum samples were assayed on CM10 chip and data were analyzed using Ciphergen Express Client software. Serum samples were thawed on ice and centrifuged at 14,000 rpm for 5 min at  $4^{\circ}\text{C}$ . A total of 20  $\mu\text{l}$  of supernatant and 30  $\mu\text{l}$  of U9-buffer (9 M urea, 2% CAHPS,

1% DTT) were added to a tube and agitated for 30 min at  $4^{\circ}\text{C}$  on a platform shaker. Subsequently, 450  $\mu\text{l}$  of Tris-HCl (50 mM pH 7.5) were added to the U9/serum mixture and agitated at  $4^{\circ}\text{C}$  for 2 min. This was followed by the addition of 150  $\mu\text{l}$  of 50 mM Tris-HCl (pH 7.5) and agitation for 5 min twice to activate the CM10 chips. Diluted samples (150  $\mu\text{l}$ ) were spotted onto a bio-processor containing the ProteinChip arrays and then agitated on a platform shaker for 60 min at room temperature. Excess serum was discarded and the chips were washed thrice with 150  $\mu\text{l}$  of Tris-HCl (50 mM, pH 7.5) and twice with 200  $\mu\text{l}$  of  $\text{dH}_2\text{O}$ , removed from the bioprocessor and air-dried. Prior to SELDI analysis, 1  $\mu\text{l}$  of a saturated solution of SPA was applied onto each chip twice and air-dried.

**Analysis of protein chip array and protein identification.** A set of different protocols was used on each spot which varied in laser intensity. Pre-processing was done on ProteinChip<sup>®</sup> software 3.1. Peaks which had less background noise were considered for further analysis after baseline subtraction. The peaks were normalized and those having  $\pm 2$  standard deviation were removed. Finally, clusters were made by expression differential matrix (EDM) within a range of 1,500-30,000 Da. Peaks having  $m/z < 1,500$  were considered as matrix noise. A threshold of 20% and signal to noise ratio and valley depth of 5 was set as first pass and 0.3% mass window and second pass of 2 was assigned. Mass accuracy was calibrated to  $< 0.1\%$  by all-in-one peptide molecular mass standard (Ciphergen Biosystems, Fremont, CA, USA).

**Biomarker validation assays.** Three independent western blot analyses were performed to validate the presence and levels of significant proteins identified through global proteome analysis. Age- and gender-matched serum samples from 10 CAD-affected and 10 CAD patients with recurrent events were used for western blot analysis. An equal amount of 50  $\mu\text{g}$  protein was loaded after protein estimation using the Bradford method. The blot was first probed for  $\beta$ -defensin-128 (sc-51670, secondary antibody SC-2033; Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) and then re-probed with histatin-3 antibody (SAB1402234; Sigma-Aldrich; secondary antibody sc-2005; Santa Cruz Biotechnology, Inc.). Finally, the same blot was probed for loading control  $\beta$ -tubulin (sc-85539; Santa Cruz Biotechnology, Inc.). Following 3 independent western blot analyses, the images were scanned in Bio-Rad molecular imaging ChemiDoc XRS plus (Bio-Rad), and bands were quantified using ImageJ software (<http://rsbweb.nih.gov/ij/index.html>) and ratio of  $\beta$ -defensin-128/ $\beta$ -tubulin and histatin-3/ $\beta$ -tubulin were calculated (Fig. 2B). ELISAs were performed for  $\beta$ -defensin-128 (CSB-EL006698HU) and histatin-3 (CSB-EL006698HU) using kits from Cusabio (Hubei, China) with covariance percentage of 1.21 and 7.04%, respectively.

**Bioinformatics and statistical analysis.** We compared the differential pattern of protein levels between the first-time CAD-affected group and the recurrent event group. Prior to the analyses, all spectra were normalized, and subjected to baseline correction and de-noising. A non-parametric t-test (Mann-Whitney) was performed to identify the most significant peaks. TagIdent software ([www.expasy.org](http://www.expasy.org)) was used to identify the proteins from their corresponding

Table I. Baseline characteristics of the study population.

Variables	First-time cases, n=85	Recurrent events, n=85	p-value
Age (years)	55.13±1.00	54.45±1.12	0.649
BMI (kg/m <sup>2</sup> )	25.83±0.416	26.95±0.486	0.080
Sex, n (%)			
Male	58 (68.2)	59 (69.4)	0.50
Female	27 (31.8)	26 (30.6)	
Hypertension, n (%)	36 (42.4)	51 (60.0)	0.031
Diabetes, n (%)	39 (45.9)	46 (54.1)	0.357
Systolic blood pressure (mmHg)	127.73±1.70	134.67±2.63	0.028
Diastolic blood pressure (mmHg)	81.47±1.003	83.11±1.067	0.266
Total cholesterol (mg/dl)	154.18±5.72	162.31±4.63	0.270
TG (mg/dl)	156.02±9.06	150.24±8.07	0.630
HDL (mg/dl)	36.76±1	35.54±1	0.391
LDL (mg/dl)	82.77±3.68	96.94±4.07	0.011

BMI, body mass index; TG, triglycerides; HDL, high-density lipoprotein, LDL, low-density lipoprotein.

m/z values (Table IIA) (9-12). Backward logistic regression, logistic regression, Cox regression analysis for hazards ratio (HR) and C-statistics were performed for the ELISA data to estimate the improvement in risk assessment of the predictive models using the two biomarkers. Using the probabilities scores through logistic regression followed by ROC analysis, we generated the area under the receiver operating characteristic curve (AUC) for 3 different models: conventional risk factors (CRFs) [age, gender, hypertension (HTN), diabetes mellitus (DM), body mass index (BMI), waist-hip ratio, smoking, total cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL) and low-density lipoprotein (LDL)] alone, the biomarkers alone, and for a combination of CRFs and biomarkers. The significance of improvement or change in the models was calculated using the DeLong method. The combined risk score was calculated for the final model of biomarkers values adjusted to CRFs. We used the following equation to calculate the combined risk score for these two biomarkers (13): combined score = ( $\beta_1 \times \beta$  - biomarker1) + ( $\beta_2 \times$  biomarker2) where,  $\beta_1$  and  $\beta_2$  denote the estimates of biomarker 1  $\beta$  coefficient ( $\beta$ -defensin-128) and biomarker 2 (histatin-3), respectively. Participants were categorized into quintiles of combined risk score with the lowest two quintiles categorized as low risk, third and fourth quintiles as intermediate, and the top quintile as high risk categories. The Kaplan-Meier curve analysis was performed for the best models for estimating the cumulative cardiovascular events in each group and plots were generated for the low and high risk groups. The analysis was performed using SPSS software (version 17.0) and R package 2.13.0.

## Results

**Patient characteristics.** Table I shows the clinical characteristics of the study participants. The mean age, gender distribution, BMI and lipid profile were comparable between the patients with first-time CAD and those with recurrent events. Hypertension and a

higher systolic blood pressure were more frequent, and LDL-c levels were higher in the latter group than in the former group of patients. The subgroup with recurrent events included 21 deaths due to cardiac arrest out of 85 subjects, while none of the patients in the first-time CAD-affected group had any event or death.

**Protein profiling.** The proteomic profile of the 85 first-time CAD-affected and 85 subjects with recurrent events using SELDI-TOF MS yielded 39 differentially expressed protein peaks [representative peak and pseudo gel (Fig. 1)], of which 16 had p-values <0.05 (Table IIA), belonging to different pathways and functions, as suggested in the SWISSPROT database ([www.expasy.org](http://www.expasy.org)). Of the 16 proteins, 4 of them were related to infection [histatin-3 (m/z 1864), defensin-5 (m/z 7040), lactotransferrin (m/z 3185) and  $\beta$ -defensin-128 (m/z 8588)]. The defensin proteins and complement C3 (m/z 2022) play an important role in immune regulatory mechanisms. A total of 5 of the other proteins, vascular endothelial growth factor A (m/z 22259), choriogonadotrophin subunit  $\beta$ -variant-1 (m/z 14680), peptide-YY (m/z 4146), growth-regulated  $\alpha$  protein (m/z 7604) and pro-neuropeptide Y (m/z 4293) were all related to growth and hormone pathways. The other members included inflammation-related proteins, interleukin-24 (m/z 18072), metalloproteinase inhibitor 4 (m/z 22466), platelet factor 4 (m/z 7742) and natriuretic peptide-B (m/z 2498), a well known CAD risk marker, cardiac phospholamban (m/z 6100) and transcription factor MAFG (m/z 17845).

Backward logistic regression analysis resulted in the identification of peaks with m/z 8588 and m/z 1864 identified as  $\beta$ -defensin-128 and histatin-3 (HTN-3), respectively, to be significantly associated with odds ratios of 7.49 (95% CI, 2.71-21.90, p-value <0.001) and 1.4 (95% CI, 1.02-2.13, p-value=0.042), respectively (Table IIB). These markers were further used for validation.

**Biomarker validation.** The two protein peaks, m/z 8588 and m/z 1864, corresponding to  $\beta$ -defensin-128 and histatin-3 were

Table II. Significant protein peaks classifying between CAD cases and subjects with recurrent CAD.

A, List of m/z protein peaks showing significant differential levels in the subjects with first-time CAD and those with recurrent events.					
Protein peaks (m/z)	Identified protein	Protein ID	m/z peak intensity in CAD affected subjects	m/z peak intensity in recurrent event subjects	p-value
1864	Histatin-3, HIS3_HUMAN	P15516	60.53±10.61	118.21±14.48	<0.001
22259	Vascular endothelial growth factor A, VEGFA_HUMAN	P15692	4.25±0.19	3.69±0.19	0.004
2022	Complement C3, CO3_HUMAN	P01024	143.75±14.44	204.44±18.34	0.006
18072	Interleukin-24, IL-24_HUMAN	Q13007	2.26±0.18	1.76±0.15	0.016
14680	Choriogonadotropin subunit β variant 1, CGB1_HUMAN	A6NKQ9	14.27±0.84	11.34±0.68	0.02
4146	Peptide YY, PYY	P10082	49.22±8.22	100.63±18.03	0.02
7604	Growth-regulated α protein, GROA_HUMAN	P09341	50.84±2.85	43.23±2.91	0.023
7040	Defensin-5, DEF5_HUMAN	Q01523	24.82±3.26	28.33±4.62	0.024
22466	Metalloproteinase inhibitor 4, TIMP4	Q99727	4.45±0.24	3.87±0.21	0.025
4293	Pro-neuropeptide Y, NPY	P01303	78.46±14.48	92.52±12.96	0.03
7742	Platelet factor 4, PLF4_HUMAN	P02776	67.07±5.81	88.58±10.14	0.037
2498	Natriuretic peptides B, ANFB_HUMAN	P16860	51.32±2.79	46.55±2.69	0.039
6100	Cardiac phospholamban, PPLA_HUMAN	P26678	22.25±2.30	25.01±2.00	0.039
3185	Lactotransferrin, TRFL_HUMAN	P02788	100.13±10.87	124.79±12.32	0.042
8588	β-defensin-128, DB128_HUMAN	Q7Z7B8	369.09±25.49	424±28.20	0.042
17845	Transcription factor MafG, MAFG_HUMAN	O15525	2.33±0.21	1.86±0.16	0.049
B, Significant peaks obtained with backward logistic regression from all the 16 m/z peaks					
m/z peak	Odds ratio	95% CI	p-value		
8588 (β-defensin-128)	7.49	2.57-21.90	<0.001		
1864 (HTN3)	1.4	1.02-2.13	0.042		
HTN3, histatin-3; DB128, β-defensin-128.					

validated using western blot analysis (Fig. 2A). Densitometric analysis (Fig. 2B) of the protein bands suggested that the levels of both proteins were higher in subjects with recurrent events than in the subjects with first-time CAD. Furthermore, ELISAs (Fig. 2C) performed for these two proteins in all the 170 subjects confirmed that both proteins were significantly differentially expressed. The mean levels of  $\beta$ -defensin-128 in the first-time CAD-affected subjects was  $0.779 \pm 0.045$  ng/ml and in the recurrent subjects was  $1.19 \pm 0.082$  ng/ml ( $p$ -value= $2.2 \times 10^{-15}$ ). Similarly, the histatin-3 level was  $2.38 \pm 0.130$   $\mu$ g/ml in the first-time CAD affected subjects and  $3.42 \pm 0.306$   $\mu$ g/ml in the recurrent subjects ( $p$ -value of 0.002).

*Association of  $\beta$ -defensin-128 and histatin-3 in CAD-affected subjects with recurrent events.* We evaluated individual and combined association of these biomarkers with recurrent events. As shown in Table III, the HR for  $\beta$ -defensin-128 was 1.715 (95% CI, 1.32-2.24;  $p$ -value= $8.14 \times 10^{-5}$ ) and that for histatin-3 was 1.133 (95% CI, 1.04-1.24;  $p$ -value=0.004).

However, when combined with the CRFs, the HRs improved to 1.729 (95% CI, 1.28-2.34;  $p$ -value= $3.51 \times 10^{-4}$ ) for  $\beta$ -defensin-128 and 1.147 (95% CI, 1.04-1.26;  $p$ -value=0.004) for histatin-3. The combined HR of  $\beta$ -defensin-128 and histatin-3 together was 1.699 (95% CI, 1.29-2.24;  $p$ -value= $1.6 \times 10^{-4}$ ) before adjustment and 1.833 (95% CI, 1.34-2.50;  $p$ -value= $1.3 \times 10^{-4}$ ) after adjustment for the conventional risk factors. The analysis of C-statistics (Fig. 3) suggested that CRFs alone yielded an AUC of 0.677 (95% CI, 0.601-0.714;  $p$ -value= $5.1 \times 10^{-5}$ ). Upon the addition of biomarkers, the AUC increased to 0.800 (95% CI, 0.702-0.844;  $p$ -value= $1.37 \times 10^{-9}$ ) and the difference was statistically significant ( $p$ -value=0.001) (Table IV) as evaluated using DeLong analysis.

The combined risk score (Table V) of  $\beta$ -defensin-128 and histatin-3 adjusted to the CRFs reclassified the subjects into 3 groups, namely low (combined risk score of <0.6256, reference tertile), intermediate (combined risk score of 0.625 to <0.956; HR, 2.22; 95% CI, 1.008-4.891;  $p$ -value=0.048) and high risk (combined risk score of  $\geq 0.956$ ; HR, 2.962; 95% CI, 1.675-

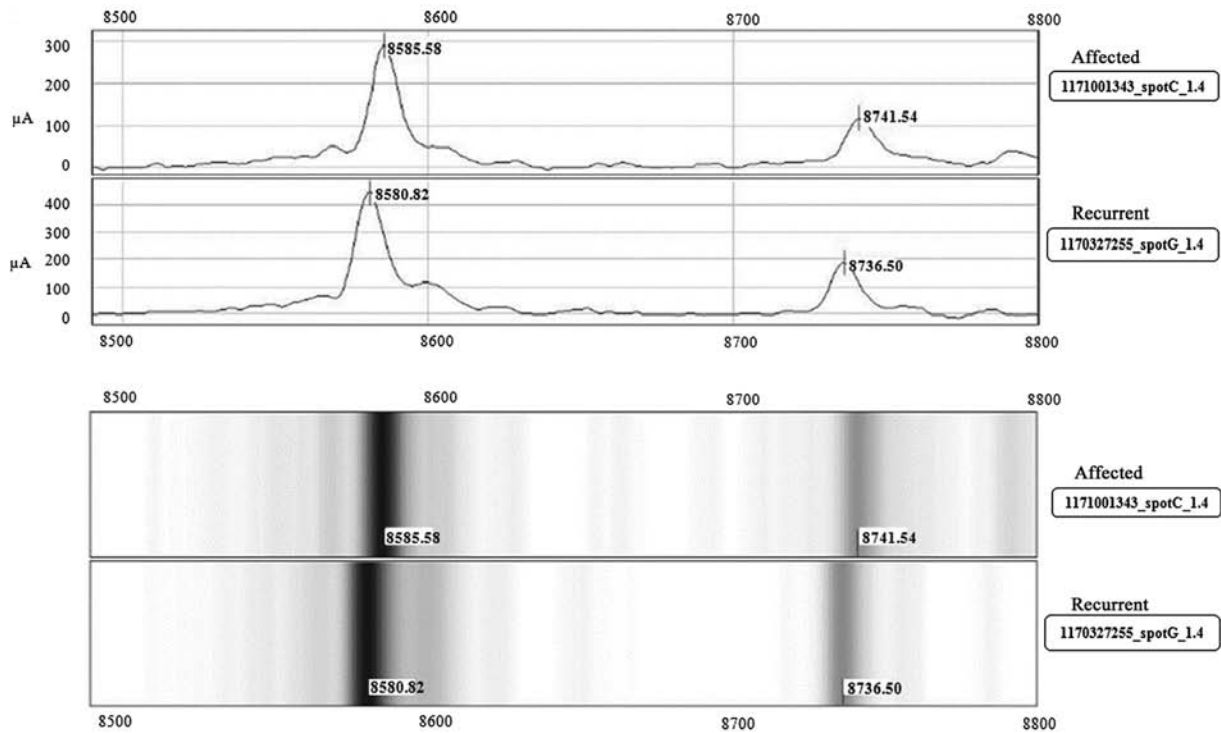


Figure 1. Representative spectra and pseudo-gel showing m/z 8588 of subjects with first-time coronary artery disease (CAD) and those with recurrent events.

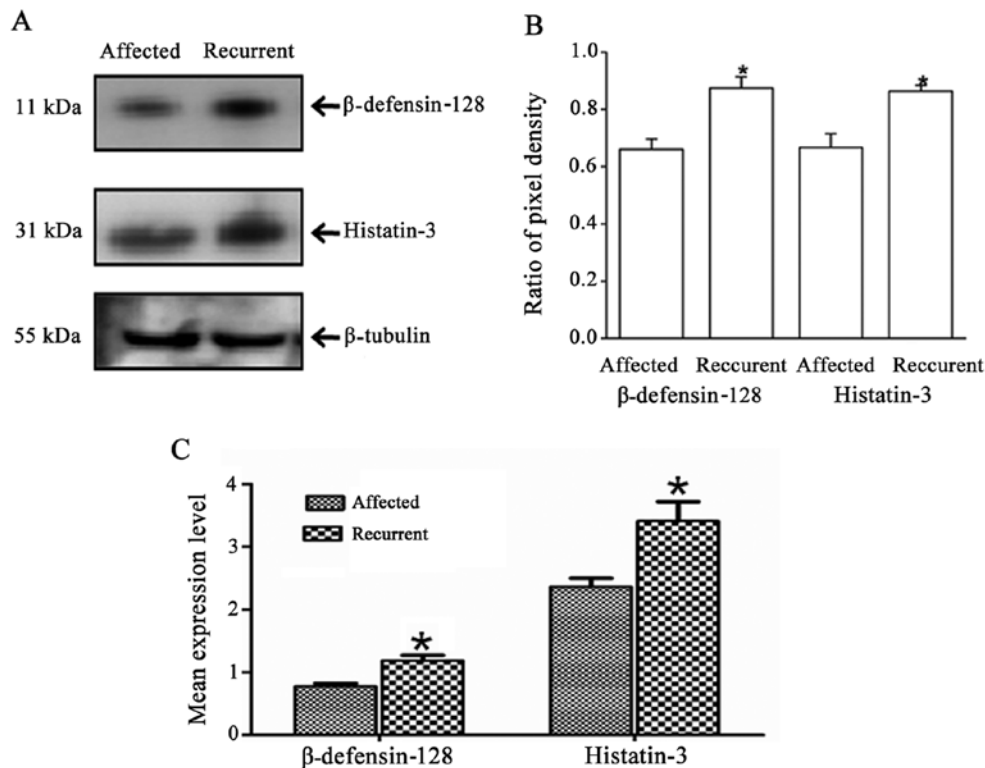


Figure 2. Validation of presence of β-defensin-128 (DB128) and histatin-3 (HTN3) in serum samples of subjects with first-time coronary artery disease (CAD) (Affected) and those with recurrent events (Recurrent). (A) Protein expression levels examined by western blot analysis. (B) Intensity of protein bands. (C) ELISA of the expression levels of β-defensin-128 (DB128) and histatin-3 (HTN3).

5.23;  $p$ -value= $1.89 \times 10^{-4}$ ). The analysis of the complete study group ( $n=170$ ) revealed that 81 subjects were at high risk of which 31 belonged to the CAD-affected group and 51 were subjects with recurrent events (13 deaths and 3 subjects with

a third reported event). Similarly, 32 subjects were reclassified into the intermediate risk group, of which 17 were from the affected group, while 15 belonged to the recurrent events group (3 deaths and 2 subjects with 3 events reported). Kaplan-

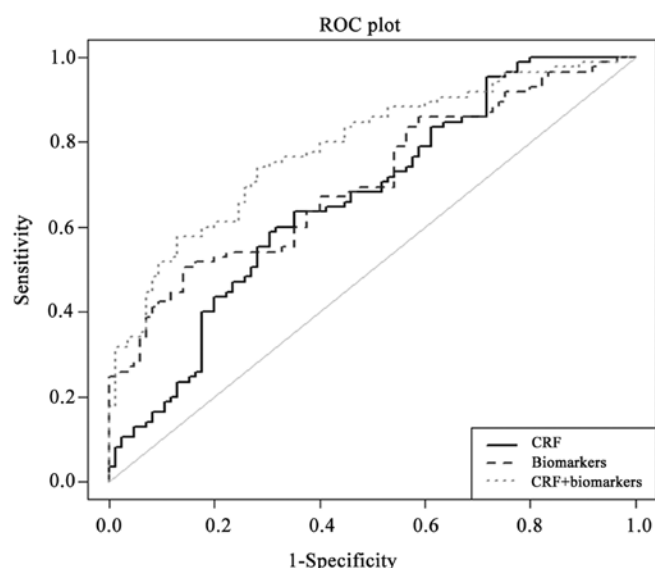


Figure 3. C-Statistic analysis of receiver operating curve showing an improvement of area under the curve for combination of CRFs and biomarkers over the CRFs alone. CRF, conventional risk factor.

Meier survival analysis for the cumulative probability of recurrent event for the model containing  $\beta$ -defensin-128 and histatin-3 adjusted to the CRFs, suggested that subjects in the high-risk group had a low survival rate as compared to those in the low-risk group (Fig. 4A). Similar analysis on the cumulative probabilities of death due to cardiovascular events or occurrence of triple events for recurrent event subjects alone also suggested that the high-risk group had a low survival rate as compared to the low-risk subjects (Fig. 4B).

## Discussion

Using comparative global proteome analysis, we identified  $\beta$ -defensin-128 and histatin-3 as novel biomarkers associated with recurrent coronary events. Recently, we also demonstrated that pathogen burden (5) and in specific CMV-NA titer is strongly associated with recurrent CAD events (6). Several other studies have focused on using putative candidate biomarkers for improving risk assessment in cases with recurrent CAD events, and these have provided a modest improvement in the risk discrimination (4). These findings suggest that there is a need to improve these discriminative models, as it is well known that subjects with established CAD have a high risk of further vascular events compared to disease-free subjects (14). In such a scenario, the use of global proteomic analysis help to unearth newer biomarkers that can substantially improve risk assessment. Global proteome analysis using SELDI-TOF MS commonly used for biomarker discovery in complex diseases has also been described with regard to network biomarkers associated with the incidence of CAD (15), major adverse cardiac events (MACE) (16) and the identification of soluble tumor necrosis factor-like weak inducer of apoptosis (sTWEAK) as a possible biomarker of subclinical atherosclerosis (17).

Of the 16 significant peaks (Table IIA), logistic regression analysis identified 2 significantly associated proteins,  $\beta$ -defensin-128 and histatin-3 (Table IIB). Furthermore,

Table III. Hazards ratio for histatin-3 and  $\beta$ -defensin-128 adjusted for conventional risk factors.

Model	Hazards ratio	95% CI	p-value
DB128	1.715	1.32-2.24	$8.14 \times 10^{-5}$
DB128 adjusted for CRF	1.729	1.28-2.34	$3.51 \times 10^{-4}$
HTN-3	1.133	1.04-1.24	0.004
HTN-3 adjusted for CRF	1.147	1.04-1.26	0.004
DB128 + HTN3	1.699	1.29-2.24	$1.60 \times 10^{-4}$
DB128 + HTN3 adjusted for CRF	1.833	1.34-2.50	$1.30 \times 10^{-4}$

HTN3, histatin-3; DB128,  $\beta$ -defensin-128.

western blot analysis (Fig. 2A and B) and ELISAs (Fig. 2C) in individual serum samples also revealed differential levels of these proteins in the 2 groups.  $\beta$ -defensins play an important role in immunoregulation (18), and have been shown to be significantly associated progressive cardiovascular disease (19). Furthermore, genetic polymorphisms and copy number variation of  $\beta$ -defensin-4 gene have been shown to be associated with ischemic stroke (20). Other studies have shown that in oral infections by spirochetes, which constitute 40% of the microflora in disease sites, are also linked to cardiovascular diseases (21,22). In our analysis,  $\beta$ -defensin-128 was shown to be associated with recurrent events, suggesting that this may be a novel biomarker belonging to the defensin family, linking infection and immunity to CAD. Similarly, gene polymorphisms in histatin-3 (23) and gene expression (24) have been shown to play a major role in oral and other infections.

Of the other proteins identified (Table IIA), we found peptides, such as natriuretic peptide-B, pro-neuropeptide-Y and peptide-YY which are associated with cardiovascular diseases (25). However, in the present study, infection-related biomarkers,  $\beta$ -defensin-128 and histatin-3 in particular seem to play a major role in the discrimination between first-time CAD-affected subjects and those with recurrent events (Tables III-V and Figs. 3 and 4). This may be due to the importance of active infection of CMV in these subjects, as our previous study using the same subset of patients demonstrated that higher titer values of CMV-NA (6) were associated with first-time CAD and to recurrent CAD (26).

Several studies have shown that chronic infection can lead to pro-atherogenic effects (27), and can lead to vascular damage (27-29). The Asian population has been earlier reported to have a higher pathogen burden, which includes CMV infection (5,6,26) and the identification of these two biomarkers suggests that they may add better value proposition in conjunction with other biomarkers. Earlier studies have reported that the combination of biomarkers, such as Nt-proBNP, albuminuria and CRP (4) or CMV-NA with IL-6 (6) has a marginal improvement in C-statistics. Therefore, the development of a multimarker approach by including  $\beta$ -defensin-128 and histatin-3, where there was a significant improvement in the AUC (Table IV), may help to better discriminate subjects

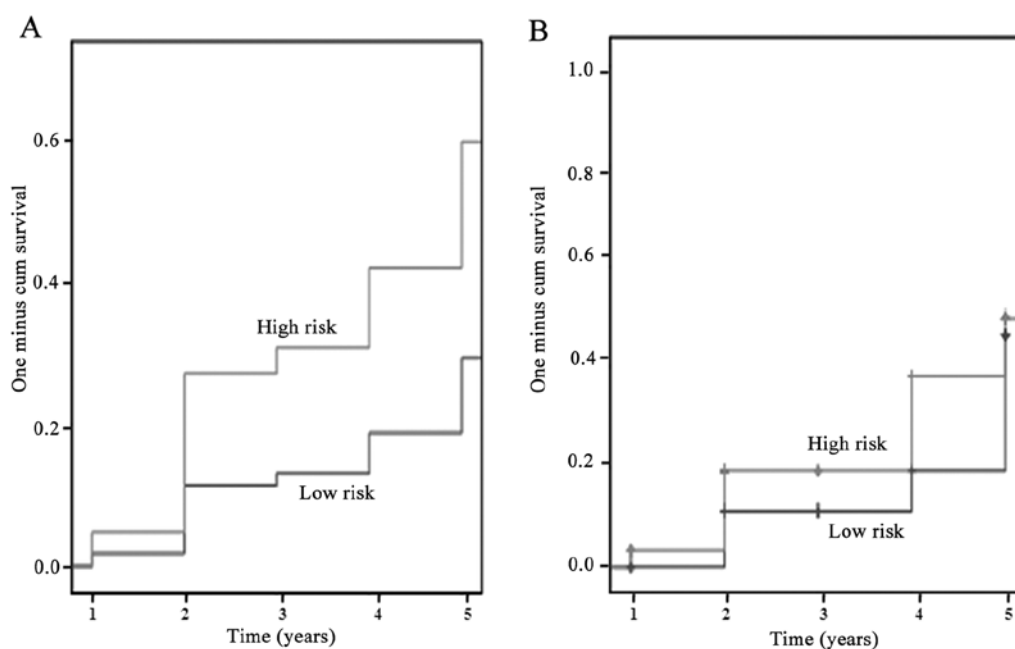


Figure 4. (A) Kaplan-Meier curve of the cumulative probability of cardiovascular recurrent events in accordance with different risk categories using a combined score of histatin-3 and  $\beta$ -defensin-128 adjusted to the conventional risk factors in a 5-year follow-up study in 85 subjects with recurrent events. (B) Kaplan-Meier curve of the cumulative probability of cardiovascular death and third event in the recurrent events subjects in low- and high-risk categories using a combined risk score of histatin-3 and  $\beta$ -defensin-128, adjusted to conventional risk factors (CRFs).

Table IV. C-statistic analysis showing improvement of the Area Under the Curve (AUC) for the combined model of biomarkers and CRF over the CRF alone.

Model	AUC	p-value	95% CI	DeLong analysis for significance (p-value) in the improvement of AUC from CRF model to CRF + biomarkers
CRF	0.677	$5.100 \times 10^{-5}$	0.601-0.714	0.001
CRF + HTN3 + DB128	0.800	$1.37 \times 10^{-9}$	0.702-0.844	

HTN3, histatin-3; DB128,  $\beta$ -defensin-128.

Table V. Combined score of histatin-3 and  $\beta$ -defensin128 adjusted for CRFs stratifying low, intermediate and high risk groups in the study population.

Category	Combined-marker score for event	HR	95% CI for Exp(B)	p-value	Total subjects reclassification n=170	Affected subjects reclassification n=85	Recurrent subjects reclassification n=85
Low	<0.6256	1.00 (Ref)	-	-	57	37	20 (6 deaths)
Intermediate	0.625 to <0.956	2.220	1.008-4.891	0.048	32	17	15 (3 deaths and 2 subjects with 3rd event)
High	$\geq 0.956$	2.962	1.675-5.23	$1.89 \times 10^{-4}$	81	31	50 (13 deaths and 3 subjects with 3rd event)

with recurrent CAD. Furthermore, the combined risk scores of  $\beta$ -defensin-128 and histatin-3 (Table V) and Kaplan-Meier curve cumulative probability analysis of cardiovascular

events (Fig. 4A) or deaths/3rd event (Fig. 4B) suggested that 68 out of the 85 subjects with recurrent events were correctly classified into the intermediate- and high-risk groups.



Understanding the limitations of the present study, further validation of these biomarkers in larger cohort of subject is needed. In order to understand the role of other proteins, protein-protein interactions network (15), pathways, regulome analysis (30) and greater dissection of all the 16 significant proteins in our data set (Table IIA) may provide a comprehensive understanding of the crosstalk that occurs among the different pathways and their collective contribution to recurrent coronary events.

In conclusion, this study identified two members of the danger-recognizing protein family,  $\beta$ -defensin-128 and histatin-3 which are associated with recurrent CAD events, and also demonstrated their potential use in discrimination of subjects.

## Acknowledgements

We are grateful for the support of the funding agencies Department of Biotechnology, Ministry of Science and Technology, Government of India (BT/01/CDE/08/07), the Tata Social Welfare Trust, India (TSWT/IG/SNB/JP/Sdm), Bharati Foundation, India, Garfield Weston Foundation, UK and Foundation Bay, UK. The sponsors did not participate in the design, conduct, sample collection, analysis and interpretation of the data or in the preparation, review or approval of the manuscript. The authors would like to thank all the investigators, staff and administrative teams and participants of IARS from Narayana Hrudayalaya, Bangalore and Asian Heart Centre, Mumbai for their contribution. The authors are grateful to the patients and their family members for participating in the study.

## References

- Sharma M and Ganguly NK: Premature coronary artery disease in Indians and its associated risk factors. *Vasc Health Risk Manag* 1: 217-225, 2005.
- Briffa TG, Hobbs MS, Tonkin A, Sanfilippo FM, Hickling S, Ridout SC and Knuiman M: Population trends of recurrent coronary heart disease event rates remain high. *Circ Cardiovasc Qual Outcomes* 4: 107-113, 2011.
- Sabatine MS, Morrow DA, de Lemos JA, Gibson CM, Murphy SA, Rifai N, McCabe C, Antman EM, Cannon CP and Braunwald E: Multimarker approach to risk stratification in non-ST elevation acute coronary syndromes: Simultaneous assessment of troponin I, C-reactive protein, and B-type natriuretic peptide. *Circulation* 105: 1760-1763, 2002.
- Shlipak MG, Ix JH, Bibbins-Domingo K, Lin F and Whooley MA: Biomarkers to predict recurrent cardiovascular disease: The Heart and Soul Study. *Am J Med* 121: 50-57, 2008.
- Mundkur LA, Rao VS, Hebbagudi S, Shanker J, Shivanandan H, Nagaraj RK and Kakkar VV: Pathogen burden, cytomegalovirus infection and inflammatory markers in the risk of premature coronary artery disease in individuals of Indian origin. *Exp Clin Cardiol* 17: 63-68, 2012.
- Mundkur LA, Shivanandan H, Hebbagudi S, Endr sz V, Varma M, Rao V, Gonczol E and Kakkar VV: Human cytomegalovirus neutralising antibodies and increased risk of coronary artery disease in Indian population. *Heart* 98: 982-987, 2012.
- Vasan RS: Biomarkers of cardiovascular disease: Molecular basis and practical considerations. *Circulation* 113: 2335-2362, 2006.
- Shanker J, Maitra A, Rao VS, Mundkur L, Dhanalakshmi B, Hebbagudi S and Kakkar VV: Rationale, design and preliminary findings of the Indian Atherosclerosis Research Study. *Indian Heart J* 62: 286-295, 2010.
- Khemika Lomthaisong, Kanchanit Boonmaleerat and Aphinya Wongpia: Proteomic study of recombinant *Escherichia coli* expressing *Beauveria bassiana* whitinase gene. *Chiang Mai J Sci* 35: 324-330, 2008.
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD and Hochstrasser DF: Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112: 531-552, 1999.
- Wilkins MR, Gasteiger E, Sanchez JC, Appel RD and Hochstrasser DF: Protein identification with sequence tags. *Curr Biol* 6: 1543-1544, 1996.
- Wilkins MR, Gasteiger E, Tonella L, Ou K, Tyler M, Sanchez JC, Gooley AA, Walsh BJ, Bairoch A, Appel RD, *et al*: Protein identification with N and C-terminal sequence tags in proteome projects. *J Mol Biol* 278: 599-608, 1998.
- Wang TJ, Gona P, Larson MG, Tofler GH, Levy D, Newton-Cheh C, Jacques PF, Rifai N, Selhub J, Robins SJ, *et al*: Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med* 355: 2631-2639, 2006.
- Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, Hailpern SM, Ho M, Howard V, Kissela B, *et al*: American Heart Association Statistics Committee and Stroke Statistics Subcommittee: Heart disease and stroke statistics - 2008 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 117: e25-e146, 2008.
- Vangala RK, Ravindran V, Kamath K, Rao VS and Sridhara H: Novel network biomarkers profile based coronary artery disease risk stratification in Asian Indians. *Adv Biomed Res* 2: 59, 2013.
- Jin G, Zhou X, Wang H, Zhao H, Cui K, Zhang XS, Chen L, Hazen SL, Li K and Wong ST: The knowledge-integrated network biomarkers discovery for major adverse cardiac events. *J Proteome Res* 7: 4013-4021, 2008.
- Blanco-Colio LM, Mart n-Ventura JL, Mu   z-Garc  a B, Orbe J, P  ramo JA, Michel JB, Ortiz A, Meilhac O and Egido J: Identification of soluble tumor necrosis factor-like weak inducer of apoptosis (sTWEAK) as a possible biomarker of subclinical atherosclerosis. *Arterioscler Thromb Vasc Biol* 27: 916-922, 2007.
- Weinberg A, Jin G, Sieg S and McCormick TS: The yin and yang of human beta-defensins in health and disease. *Front Immunol* 3: 294, 2012.
- Vordenb  umen S, Sander O, Bleck E, Schneider M and Fischer-Betz R: Cardiovascular disease and serum defensin levels in systemic lupus erythematosus. *Clin Exp Rheumatol* 30: 364-370, 2012.
- Tiszlavicz Z, Somogyv  ri F, Szolnoki Z, Sztr  ha LK, N  meth B, V  csei L and M  ndi Y: Genetic polymorphisms of human  $\beta$ -defensins in patients with ischemic stroke. *Acta Neurol Scand* 126: 109-115, 2012.
- Brisette CA and Lukehart SA: *Treponema denticola* is resistant to human beta-defensins. *Infect Immun* 70: 3982-3984, 2002.
- de Koning J, Hoogkamp-Korstanje JA, van der Linde MR and Crijns HJ: Demonstration of spirochetes in cardiac biopsies of patients with Lyme disease. *J Infect Dis* 160: 150-153, 1989.
- Fujigaki Y, Imamura Y, Oomori Y, Ouryouji K, Miyazawa H and Wang PL: Polymorphism of salivary histatin gene and periodontal disease in the Japanese population. *J Int Acad Periodontol* 11: 220-225, 2009.
- Burgener A, Mogk K, Westmacott G, Plummer F, Ball B, Broliden K and Hasselrot K: Salivary basic proline-rich proteins are elevated in HIV-exposed seronegative men who have sex with men. *AIDS* 26: 1857-1867, 2012.
- Hung CC, Pirie F, Luan J, Lank E, Motala A, Yeo GS, Keogh JM, Wareham NJ, O'Rahilly S and Farooqi IS: Studies of the peptide YY and neuropeptide Y2 receptor genes in relation to human obesity and obesity-related traits. *Diabetes* 53: 2461-2466, 2004.
- Ji YN, An L, Zhan P and Chen XH: Cytomegalovirus infection and coronary heart disease risk: A meta-analysis. *Mol Biol Rep* 39: 6537-6546, 2012.
- Epstein SE, Zhu J, Najafi AH and Burnett MS: Insights into the role of infection in atherogenesis and in plaque rupture. *Circulation* 119: 3133-3141, 2009.
- Albert LJ and Inman RD: Molecular mimicry and autoimmunity. *N Engl J Med* 341: 2068-2074, 1999.
- Epstein SE: The multiple mechanisms by which infection may contribute to atherosclerosis development and course. *Circ Res* 90: 2-4, 2002.
- Vangala RK, Ravindran V, Ghatge M, Shanker J, Arvind P, Bindu H, Shekar M and Rao VS: Integrative bioinformatics analysis of genomic and proteomic approaches to understand the transcriptional regulatory program in coronary artery disease pathways. *PLoS One* 8: e57193, 2013.