RESEARCH ARTICLE

Understanding gene expression in coronary artery disease through global profiling, network analysis and independent validation of key candidate genes

PRATHIMA ARVIND, SHANKER JAYASHREE*, SRIKARTHIKA JAMBUNATHAN, JINY NAIR and VIJAY V. KAKKAR

Mary and Garry Weston Functional Genomics Unit, Thrombosis Research Institute, Bengaluru 560 099, India

Abstract

Molecular mechanism underlying the patho-physiology of coronary artery disease (CAD) is complex. We used global expression profiling combined with analysis of biological network to dissect out potential genes and pathways associated with CAD in a representative case–control Asian Indian cohort. We initially performed blood transcriptomics profiling in 20 subjects, including 10 CAD patients and 10 healthy controls on the Agilent microarray platform. Data was analysed with Gene Spring Gx12.5, followed by network analysis using David v 6.7 and Reactome databases. The most significant differentially expressed genes from microarray were independently validated by real time PCR in 97 cases and 97 controls. A total of 190 gene transcripts showed significant differential expression (fold change >2, $P<0.05$) between the cases and the controls of which 142 genes were upregulated and 48 genes were downregulated. Genes associated with inflammation, immune response, cell regulation, proliferation and apoptotic pathways were enriched, while inflammatory and immune response genes were displayed as hubs in the network, having greater number of interactions with the neighbouring genes. Expression of *EGR1/2/3*, *IL8*, *CXCL1*, *PTGS2*, *CD69*, *IFNG*, *FASLG*, *CCL4*, *CDC42*, *DDX58*, *NFKBID* and *NR4A2* genes were independently validated; *EGR1/2/3* and *IL8* showed >8-fold higher expression in cases relative to the controls implying their important role in CAD. In conclusion, global gene expression profiling combined with network analysis can help in identifying key genes and pathways for CAD.

[Arvind P., Jayashree S., Jambunathan S., Nair J. and Kakkar V. V. 2015 Understanding gene expression in coronary artery disease through global profiling, network analysis and independent validation of key candidate genes. *J. Genet.* **94**, 601–610]

Introduction

Coronary artery disease (CAD) is a leading cause of morbidity and mortality (Koerselman *et al*. 2003) and can be attributed to the complex interactions between genetic susceptibility factors and environmental risk factors (Vinukonda *et al*. 2009). Individual genetic markers are unlikely to fully explain the complex aetiopathology of CAD whereas profiling of transcript abundance in blood can provide a snap shot of the complex genetic interactions underlying disease pathogenesis and promote a better understanding of the genetic basis of CAD (Holdt *et al.* 2010). There is evidence to show that the gene expression pattern in peripheral blood samples do mirror the gene expression changes that occur in the atherosclerotic vascular wall (Sinnaeve *et al.* 2009).

Global gene expression profiling of blood and tissue samples from patients with atherosclerosis constitutes an important landmark for discovering novel atherothrombotic genes and prognostic indicators for CAD (Ma and Liew 2003; Sinnaeve *et al*. 2009). In this regard, Dahl *et al*. (2007) identified visfatin, an important hormone secreted by the adipose tissue, to be highly expressed in the carotid plaque microarray and verified its role as an inflammatory mediator of plaque destabilization. Through an integrated approach that combined microarray and protein network analysis, transforming growth factor beta receptor 1 (*TGFBR1*) gene was identified as a new prognostic biomarker for predicting left ventricle remodelling following acute myocardial infarction (Devaux *et al*. 2011). As such, identification of differentially expressed CAD genes can lead to putative biomarkers for early disease detection and provide deeper insights into its pathophysiology.

Asian Indians have a higher predisposition to CAD as compared to any other global population. The disease is characterized by premature onset, multivessel involvement and higher mortality rates (Begom and Singh 1995; Tillin *et al*. 2008), compounded by distinct genetic architecture (Indian Genome Consortium 2005). While there are published reports

*For correspondence. E-mail: Jayashreeshanker@triindia.org.in.

Keywords. coronary artery disease; gene expression; microarray; candidate genes; biological network; Asian Indians.

on a number of CAD genes in Asian Indians (Maitra *et al*. 2008), to our knowledge there are no studies on microarray-based transcriptomics analysis in this population. We performed whole-genome expression profiling of peripheral blood samples in CAD patients, age and gender matched controls, followed by network analysis of differentially expressed genes and finally, validation of candidate genes, with an aim to identify novel atherothrombotic genes that may be of particular relevance to this 'high risk' population.

Material and methods

Study population

The test cohort of 20 subjects included 10 CAD patients (cases) and 10 matched controls, while the validation cohort included 97 cases and 97 age and gender matched controls. All subjects were selected from the Indian Atherosclerosis Research Study (IARS), which is an epidemiological study on Asian Indians established with the objective of understanding the contribution of traditional and emerging risk factors of premature CAD in this population. An overview of the IARS design has been previously published (Shanker *et al.* 2010). The study has been approved by the ethics committee of the Thrombosis Research Institute (Study Number IARS-TRI 101). All participants gave their informed written consent. The IARS has been designed according to the guidelines of the Indian Council of Medical Research for undertaking research on human subjects (Kumar 2006) and the study has been carried out according to the principles expressed in the Declaration of Helsinki.

CAD patients were selected from among those who were admitted for coronary artery bypass surgery, between May 2010 and May 2011 at Narayana Institute of Cardiac Sciences, Bengaluru, India. Inclusion criteria for the cases were angiographic evidence for presence of CAD, with $\geq$70% stenosis in any one major epicardial artery or $\geq$50% in two or more epicardial arteries and treatment by either percutaneous transluminal coronary angioplasty or bypass graft. Age at CAD onset was $\leq$60 years for men and $\leq$65 years for women. All cases had a strong family history of cardiovascular disease. Relevant medical information was obtained from laboratory test report, electrocardiogram and echocardiography available in the hospital record. We matched 10 cases and 10 controls for presence/absence of hypertension and diabetes mellitus for the microarray study. Control subjects were clinically asymptomatic for CAD, showed normal electrocardiogram, did not have family history of cardiovascular disease and were enrolled from the same geographical area as the probands.

Laboratory assays

Overnight fasting blood samples were collected from study participants. Aliquots of serum and plasma were stored at −80°C and used for biochemical analysis. Fresh whole blood EDTA samples were collected in ice and processed immediately for extraction of total RNA. Serum triglycerides (TG) and total cholesterol (TC) (Randox Laboratories, Antrim, UK) were estimated by standard enzymatic analysis on Cobas-Fara II Clinical Chemistry Auto analyzer (F. Hoffman La Roche Ltd, Switzerland). High density lipoprotein-cholesterol (HDL-c) concentrations were estimated after precipitating the nonHDL fractions with a mixture of 2.4-mmol/L phosphotungstic acid and 39 mmol/L magnesium chloride (Bayer Diagnostics, Baroda, India). Precipitating reagent was obtained from Bayer Diagnostics (India). Plasma low density lipoprotein-cholesterol (LDL-c) values were calculated using Friedwald's formula (Friedewald *et al.* 1972).

Isolation of total RNA

Total RNA was extracted using QIAamp RNA Blood mini kit (Qiagen, Valencia, USA) following manufacturer recommended protocol. RNA integrity number was calculated using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA) and RNA was quantified on NanoDrop ND-1000 Spectrophotometer (NanoDrop, Wilmington, USA).

Microarray experimental protocol

RNA samples were hybridized to G4851A, human genome G3 Hmn GE 8X60K microarray slides (Agilent Technologies) for one colour gene expression profiling using manufacturer recommended protocol. Briefly, cDNA was synthesized from 200 ng of total RNA, transcribed to cRNA with incorporation of Cy3-dUTP during the transcription process. Labelled probes were purified using RNAeasy mini kit (Qiagen) and concentrated to desired volume. cRNA was fragmented and hybridized to the microarray slide, placed in sureHyb chamber and incubated overnight at 65°C for 17 h in the hybridization oven. Following day, the slides were rinsed with wash buffer 1 and 2 to remove unhybridized fragments, air dried and scanned using Agilent microarray scanner. Data was extracted with Feature extraction software ver. 10.7.3.1. The dataset included for the study has been deposited in NCBI's Gene Expression Omnibus (Edgar *et al*. 2002) and is accessible through GEO series accession number GSE 42148 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42148).

Microarray data analysis

Data processing and statistical analysis: The raw data was further analysed with Gene Spring Gx 12.5 software (Agilent Technologies). All samples were normalized at 75th percentile, base line was set to median and the probes were filtered based on their intensities. Significant difference in gene expression levels between cases and controls was estimated using moderated *t*-test. A correction for multiple testing was performed using Benjamin Hochberg false discovery rate (FDR) function. Hierarchical clustering was performed with the probe list across the cases and controls using Euclidian distance similarities in Ward's method linkage.

Functional analysis: Pathway analysis was performed using bioinformatics tools like Reactome (Matthews *et al.* 2009) (http://www.reactome.org) and DAVID ver. 6.7 (database for annotation, visualization and integrated discovery) (Huang da *et al.* 2009a, b) (http://david.abcc.ncifcrf.gov/). These bioinformatics resources provide functional interpretation of the large list of genes derived from microarray experiment. Gene relationship and interaction was extracted using Natural language processing (NLP) algorithms in Pathway Architect software available in Gene Spring ver. 12.5 (Agilent technologies). This tool extracts interactions for a given list of genes from Pubmed abstracts with text mining tools as well as incorporates new significant molecules into the network. A network with direct interactions in terms of binding, expression, metabolism, promoter binding, protein modification, regulation and transport was constructed using the significant differentially expressed genes.

Independent validation of candidate genes

Twenty-five putative candidate genes showing high differential expression between cases and controls in the microarray data were selected for validation (table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/). Of the 25 genes, expression analysis for 10 genes was undertaken in 97 cases and 97 age and gender matched controls. Gene expression levels were measured using Taqman assay on ABI 7900 real-time PCR (Applied Biosystems, USA). For the remaining 15 genes, the expression analysis was carried out in 50 cases and 50 controls (subset of the validation dataset) using KAPA SYBR FAST qPCR kit (KAPA Biosystems, USA) including cDNA specific real-time PCR assays. All assays were set up in duplicates and performed as per the manufacturer's recommended protocol. Real-time data was analysed using Sequence Detection Software (SDS) ver. 2.3 (Applied Biosystems). Relative changes in mRNA expression levels between cases and controls were calculated using RQ manager (Applied Biosystems) for TaqMan gene expression data, while comparative Ct method was used for SYBR Green expression data (Livak and Schmittgen 2001).

Estimation of plasma protein levels using fluorescence-activated cell sorter (FACS)

Two key genes that were significantly differentially expressed in the cases as compared to the controls were selected for measuring the corresponding plasma protein levels. CBA Flex Set, a multiplex Cytometric bead array kit was purchased

Table 1. Clinical characteristics of study participants.

	Test cohort			Validation cohort		
Clinical factor	Case (n =10)	Control (n =10)	P value	Case (n =97)	Control (n =97)	P value
Age (years)	48.70±0.761	50.90±0.900	0.779	50.13±0.651	49.88±0.651	0.780
Age at onset (years)	46.70±1.430			-48.56±0.740		
Male N (%)	10	10		-82(84.5)	89(91.7)	
BMI (kg/m^2)	24.26±0.87	23.98±1.603	0.878	24.82±0.394	24.43±0.392	0.491
Waist circumference (cm)	87.90±1.99	87.10±4.100	0.863	89.51±0.899	91.13±0.904	0.205
Hip circumference (cm)	88.20±1.610	88.99±3.620	0.863	89.65±0.761	92.16±0.765	**0.021**
Waist/hip ratio (cm)	0.99±0.016	0.97±0.029	0.709	0.99±0.006	0.98±0.006	0.287
Laboratory studies						
TC (mg/dL)	129.11±8.200	154.30±7.949	0.052	129±3.89	179±3.87	**1.86*10^{-16}**
TG (mg/dL)	149.89±17.70	190.60±27.06	0.227	154.75±9.11	194.71±9.06	0.002
HDL-c (mg/dL)	37.22±1.498	33.70±1.814	0.072	31.02±1.028	37.84±1.023	**5.02*10^{-6}**
LDL-c (mg/dL)	61.91±6.17	83.480±9.857	0.084	68.30±3.29	103.87±3.35	**1.79*10^{-12}**
Medical history						
Smoking N (%)	5 (50.0)	5 (50.0)	0.672	59(62.1)	25 (26.3)	**5.50*10^{-7}**
Hypertension N (%)	4 (40.0)	2 (20.0)	.314	38(39.6)	11 (11.5)	**5.76*10^{-6}**
Diabetes mellitus N (%)	4 (40.0)	4 (40.0)	0.675	49(51)	17 (17.7)	**8.87*10^{-7}**
FBS mg/L	104.00±19.67	115±21.310	0.705	120.97±4.39	102.31±4.23	0.003
Medications						
Statin N (%)	9(90.0)			53(56.7)		
Beta blocker N (%)	7(70)			51(52.6)	1 (1)	**1.89*10^{-10}**
Calcium channel						
Blocker N (%)	1(10)	1(10)	0.901	6(6.2)		
ACE inhibitor N (%)	3(30)	1(10)	0.053	24(24.7)		
Antiplatelet N (%)	5(50)			48(49.5)		
Hypoglycemic agents N (%)	2(20)	4(40)	0.239	8(8.2)		
Nitrate N (%)	8(80)			41(42.13)		

Continuous variables are expressed as mean ± standard error.
BMI, body mass index; FBS, fasting blood sugar; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol; TC, total cholesterol; TG, triglycerides. Significant P value are in bold.

from BD Bioscience (BD Bioscience, San Diego, USA). The experiment was performed on BD FACSCanto II flow Cytometer (BD Bioscience). Standard curves were generated for each cytokine using the reference cytokine concentrations supplied by the manufacturer. Raw data (mean fluorescent intensity) was analysed with FCAP Array ver. 3 software to obtain protein concentrations in the individual samples.

Statistical analysis

Routine statistical analysis was performed with SPSS ver. 17.0 software (SPSS, Chicago, USA). Values are expressed as mean±SE for all continuous variables. Student *t*-test, univariate and multivariate analyses were used to calculate the mean differences in quantitative traits between the cases and controls. A P value of ≤ 0.05 was considered statistically significant. Correlation in gene expression levels was estimated using Pearson's correlation test. Age, gender and statins were treated as covariates and appropriately adjusted during analysis. The 'power of the study' was calculated using a web-based tool (Lee and Whitmore 2002) (http://sph.umd.edu/epib/faculty/mltlee/web-front-r.html) designed for microarray study. Estimation of 'power of the study' will determine the minimum sample size required to detect a significant effect on the outcome.

Results

Clinical characteristics of study participants

The clinical profile of study participants included in the microarray and validation study is provided in table 1. All participants were males with mean age of 48.70±0.76 years in cases and 50.90±0.90 years in controls. Frequency of diabetes, hypertension, smoking and anthropometric measurements were comparable between both the groups. In the validation cohort, there were around 85% male and 15% female in cases, and 92% male and 8% female in the controls. The mean age was 50.13±0.65 years in cases and 49.88±0.65 years in controls. Classical risk factors were more frequent among the cases. Anthropometric measurements were comparable between the two groups. Lipid levels were lower in the cases than in the controls, which might be attributed to the usage of statins. The average age at CAD onset was around 48 years both in the test and the validation cohort.

Differential gene expression in cases compared to controls

The global gene expression profile was available for 10 cases and 10 controls. After initial filtering of data, 190 annotated probe sets (genes) were differentially expressed with >2 fold, $P \leq 0.05$, of which 142 were upregulated and 48 were downregulated in cases than in the controls (table 2 in electronic supplementary material). Hierarchical clustering was performed to identify genes showing significant coexpression which segregated into two main clusters (figure 1). The larger cluster comprised of 142 upregulated genes that was sub divided into highly upregulated genes (in red colour) (≥2.5 fold change) having fold change ranging from >2.6 up to 50 fold (EGR genes, IL8, CDC42 etc.), and modestly upregulated genes showed between 2.0 and 2.5 fold differential expression; the smaller cluster contained 48 genes that were downregulated in the cases. The upregulated gene cluster included those associated with early growth response (e.g. *EGR3, AREG*), inflammation and immune response (e.g. *GOS2*, *IL8*, *CXCL1*, *TAGAP*, *PTPRJ*), cell regulation (e.g. *ID1*, *RGS1*, *NUDCD3*), transcription factor (e.g. *FOSB*, *NR4A2*), metabolic process (e.g. *SLC2A3, SLC11A1, C2CD3*) and diabetes pathways (e.g. *IRS2*, *SFN*,

Table 2. Pathway enrichment analysis using GeneSpring ver. 12.5, DAVID ver. 6.7 and Reactome.

Pathway name	*P* value
Upregulated pathways	
Receptor *CXCR2* binds ligand *CXCL1* to 7 (*CXCL1*, *IL8*, *CXCL6*)	0.0001
TGF beta signalling pathway (*CDC42*, *FOSB*, *JUND*, *NEDD9*)	0.0006
Receptor *CXCR1* binds *CXCL6* and *CXCL8* ligands (*IL8*, *CXCL6*)	0.0002
Activation of WASP/N-WASP by WIP family and SH3 domain proteins (*WIPF1*, *CDC42*)	0.001
MAPKinase pathway (*ABL1*, *CDC42*, *JUND*, *MAP3K13*)	0.005
EGF- EGFR signalling pathway (*ABL1*, *CDC42*, *JUND*, *FOSB*)	0.006
Chemokine receptors bind chemokines (*CXCL1*, *IL8*, *CXCL6*)	0.007
TCR signalling pathway (*CDC42*, *MAP3K13*, *CD83*)	0.010
Corticotropin-releasing hormone (*NR4A2*, *JUND*, *FOSB*)	0.011
TSLP signalling pathway (*IL8*, *RPS6*)	0.016
CDO in myogenesis (*ABL1*, *CDC42*)	0.019
IL-3 signalling pathway (*IL8*, *CD69*)	0.022
Leptin signalling pathway (*RPS6*, *CDC42*)	0.038
Insulin signalling (*SGK*, *IRS2*, *MAP3K13*)	0.047
Downregulated pathway	
Generic transcription pathway (e.g. *ZNF436*, *ZNF100*)	0.002
Gene expression (e.g. *ZNF436*, *ZNF100*, *ZNF614*)	0.002
Nucleotide GPCRs (*P2RY6*)	0.019

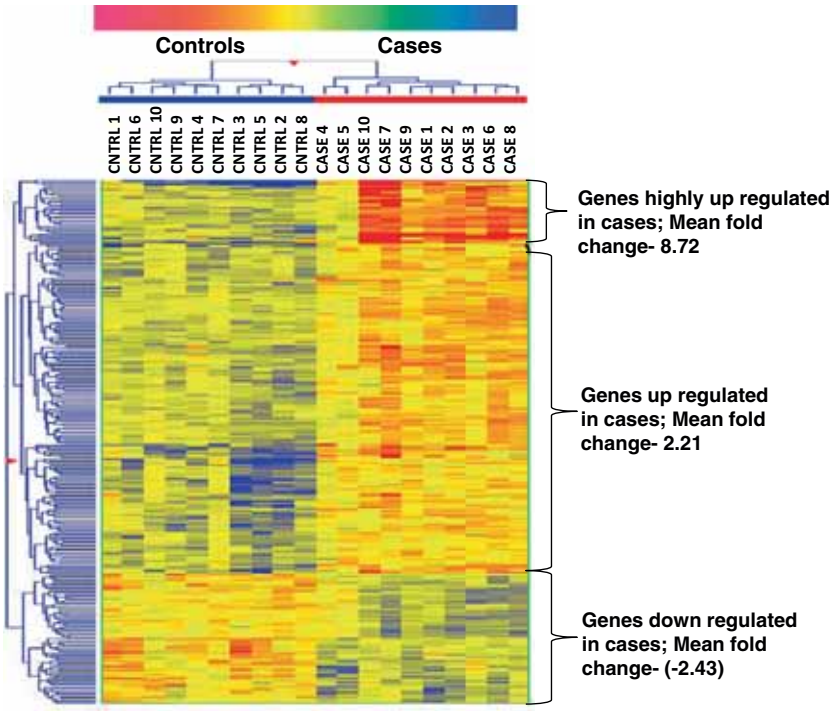

Figure 1. Heat map depicting differentially expressed genes analysed by hierarchical clustering method. The genes were classified into three major clusters based on the following parameters: Wald's linking and Euclidean distance. Samples are displayed in columns and genes in rows. Normalized gene expression is represented across each row, with red indicating higher values and blue indicating lower values.

EXOC7), while the downregulated gene cluster were associated with cell regulation (e.g. *FGF5, P2RY6*), metabolic process (*LCMT2*) etc.

Pathway analysis and gene network analysis

In pathway analysis, classical pathway databases like Kyoto Encyclopedia of Genes and Genomes (KEGG) in DAVID ver. 6.7 and other pathway analysis tools such as Reactome and GeneSpring GX ver. 12 showed maximum number of significant pathways. Table 2 lists the significant pathways identified for the upregulated and downregulated genes. Immune and inflammatory response, cell signalling and cell proliferation related genes comprised the most significant pathways. The cell division cycle 42 (*CDC42*), FBJ murine osteosarcoma viral oncogene homolog B (FOSB), jun D proto-oncogene (*JUND*) and interleukin 8 (*IL8*) genes were involved in multiple pathways. On the other hand, genes encoding for the zinc finger proteins (*ZNF436*, *ZNF100*) were downregulated in the cases. Among the differentially expressed genes, *IL8, FOSB, JUND* and *CDC42* appeared to be the node genes, displaying maximum interactions with other genes in the network (figure 2). Interestingly, *FOSB* appeared to regulate the transcription and promoter binding activity of *IL8*, while *JUND* regulated *IL8* and *IL8* in turn regulated *CDC42*). *FOSB* and *JUND* were connected by E74-like factor 3 (*ELF3*) connectivity proteins. The downregulated genes did not show any distinct interaction pattern.

Validation of microarray data

Gene expression by TaqMan assay: The chemokine (C-X-C motif) ligand 1 (*CXCL1*), early growth response 3 (*EGR3*), *IL8*, prostaglandin-endoperoxide synthase 2 (*PTGS2*), pentraxin 3 (*PTX3*), and *CD 69* molecule (*CD69*) showed significant higher expression while interferon gamma (*IFNG*) and fas ligand (*FASLG*) showed relatively lower expression in the cases than in the controls ($P \leq 0.05$). After adjusting for age, gender and statin usage, all genes except *CD69* retained statistical significance (figure 3).

IL8 and *CXCL1* that mediate inflammatory response showed strong correlation with each other ($r = 0.845$, $P < 0.0001$) and also with *PTGS2* and *PTX3* genes, that are known regulators of inflammation and mitogenesis ($r = 0.40$–64, $P = 0.0001$). Interestingly, *CD69* that induces activation of T-lymphocytes and *EGR3*, which is responsible for early growth and development of muscle, lymphocyte, endothelial cell growth and migration showed strong correlation with inflammatory genes, *CXCL1*, *IL8*, *PTGS2* and *PTX3*, ($r = 0.32$–0.62, $P = 0.0001$). Further, *CD69* expression correlated with matrix metalloproteinase 8 (MMP8) and oxidized low density lipoprotein receptor 1 (*OLR1*) ($r = 0.36$–0.38, $P = 0.0001$). Both *IFNG* (an antiinflammatory cytokine) and *FASLG* (apoptotic trigger

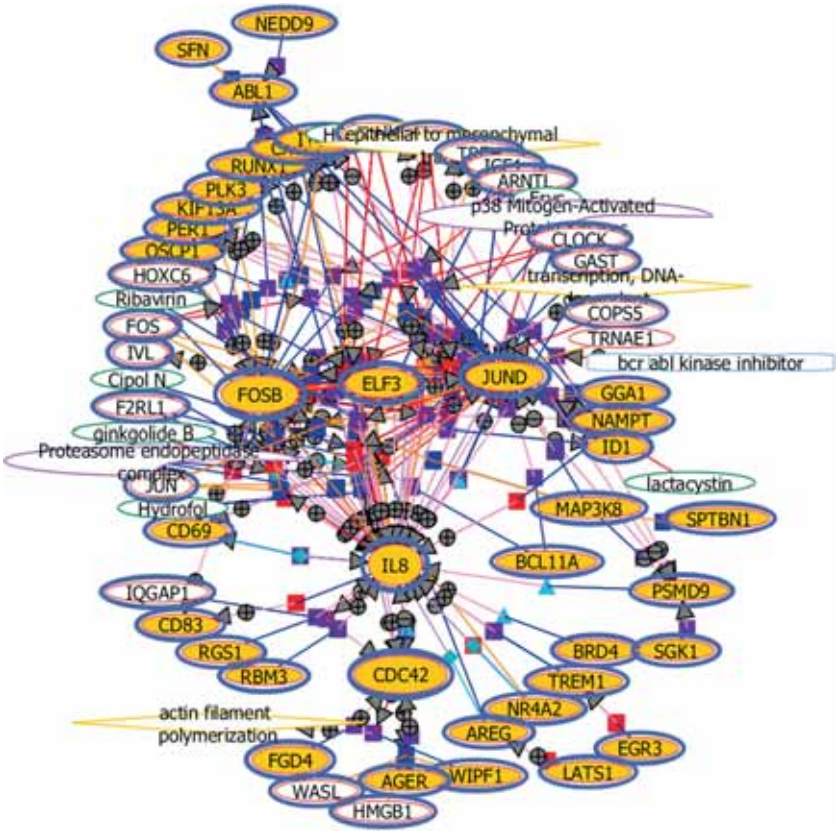

Figure 2. NLP analysis of upregulated genes in CAD. Network displayed the interaction between seed gene (yellow) and its interacting partners (white). IL8 appears to be a nodal gene showing strong interactions with most other seed genes and other genes pulled down in the network.

of lymphocytes) gene transcripts showed strong correlation with each other ($r = -0.56$, $P = 0.0001$). *FASLG* showed negative correlation with *CXCL1*, *IL8*, *EGR3*, *PTGS2* and *PTX3* ($r = -0.19$ to 0.37, $P < 0.01$). *MMP8* showed strong correlation only with *OLR1* ($r = 0.68$, $P = 0.0001$).

Gene expression by SYBR Green method: The chemokine (C–C motif) ligand 4 (*CCL4*), *CDC42*, DEAD (Asp-Glu-Ala-Asp) box polypeptide 58 (*DDX58*), early growth response genes (*EGR1/2*), nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, delta (*NFKBID*) and nuclear

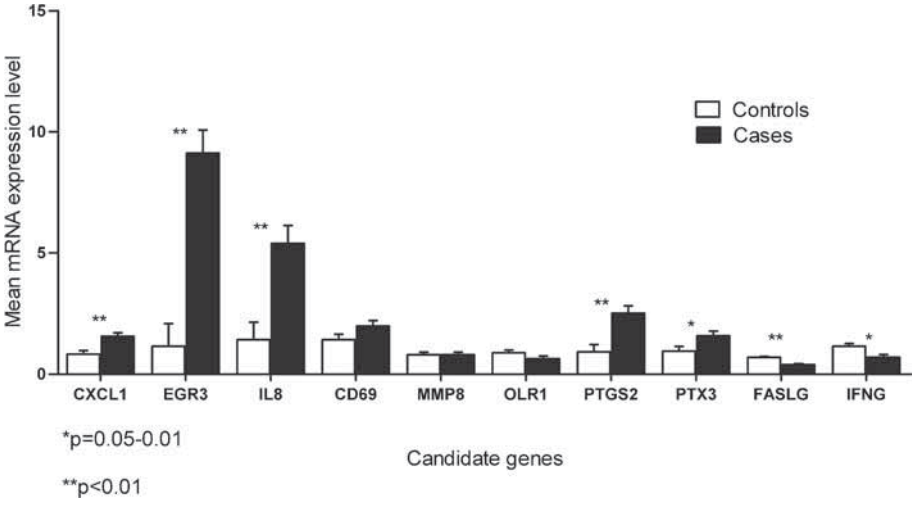

Figure 3. Mean expression levels of 10 candidate genes in cases and controls, adjusted for age, gender and statins. Data presented as standard error mean (SEM).

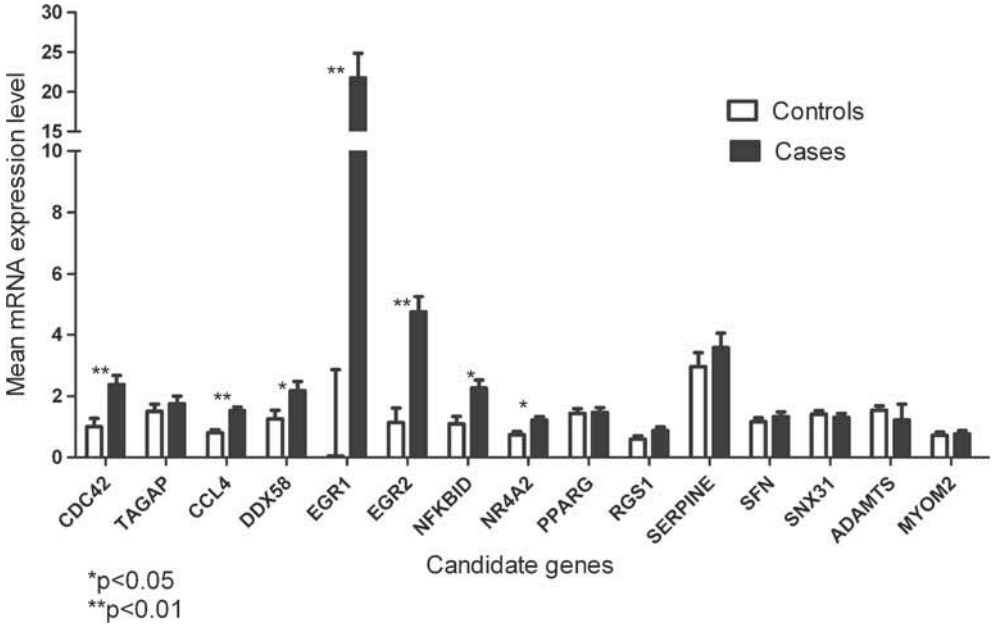

Figure 4. Mean expression levels of 15 candidate genes in cases and controls, mean expression levels adjusted for age, gender and statins. Data presented as standard error mean (SEM).

receptor subfamily 4, group A, member 2 (*NR4A2*) were significantly upregulated ($P \leq 0.05$) in cases as compared to controls after adjusting for age, gender and statin usage. The remaining genes did not show any significant difference in expression between the cases and controls (figure 4). The significant differentially expressed genes in the validation dataset and their functional impact on CAD is summarized in table 3 in electronic supplementary material.

In correlation analysis after adjusting for age and gender, *CDC42* showed positive correlation with many genes such as *CCL4, DDX58, EGR1, EGR2*, regulator of G-protein signalling 1 (*RGS1*) ($r = 0.381$–0.690, $P = \leq 0.05$) and negative correlation with A disintegrin and metalloproteinase with thrombospondin motifs (*ADAMTS*), myomesin 2 (*MYOM2*) and sorting nexin 31 (*SNX31*) ($r = -0.414$ to -0.274, $P = \leq 0.05$). *CCL4* showed correlation with most of the genes, *EGR1, EGR2, NR42A2, RGS1* and T-cell activation RhoGTPase activating protein (*TAGAP*)($r = 0.201$–0.519, $P = \leq 0.05$). *EGR1* and *EGR2* showed high correlation with each other ($r = 0.695$, $P = 2.00 \times 10^{-9}$). *EGR1* alone correlated with *DDXA58, TAGAP* and *RGS1* ($r = 0.299$–0.300, $P = \leq 0.05$) and showed negative correlation with *SNX31* ($r = -0.260$, $P = 0.05$) while *EGR2* expression correlated with *NR4A2* and *RGS1* ($r = 0.338$–0.519, ≤ 0.01). Expression *SFN* and *SNX31* genes showed correlation with each other ($r = 0.285$, $P = 0.031$) and both correlated with *ADAMTS* ($r = 0.423$–0.674. $P = \leq 0.01$). Only *SFN* showed correlation with *NR4A2* and *PPARG* ($r = 0.380$–0.393, $P = \leq 0.005$) while *SNX31* expression correlated with *MYOM2* and *NFKBID* ($r = 0.318$–0.376, $P = \leq 0.05$). *SERPINE* showed correlation with *DDX58* and *TAGAP.*

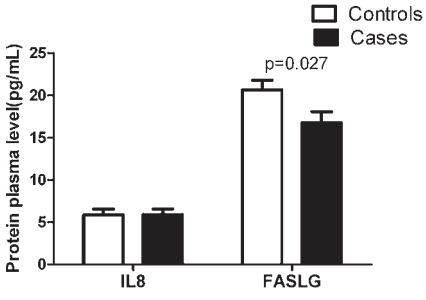

Figure 5. Mean plasma level of IL8 and FASLG proteins in cases and controls. Data presented as standard error mean (SEM).

Estimation of plasma protein levels

Plasma levels of protein products of two key genes, *IL8* and *FASLG,* were measured by CBA kit method in 50 cases and 50 controls. *FASLG* showed significantly lower levels in cases relative to the controls while the difference in *IL8* protein levels between the cases and controls was not statistically significant (figure 5).

Discussion

We analysed the whole-genome expression profile in peripheral blood sample with the primary aim of determining putative candidate genes and pathways important for CAD development in Asian Indians. Further, gene enrichment analysis and pathway analysis using curated database like Reactome and analytical tools like DAVID and Gene Spring helped in identifying a number of canonical pathways for CAD. The most significantly enriched pathways were inflammation and immune response, *TGFB* signalling pathway, *IL-3* signalling,

T-cell signalling and chemokine signalling and cell proliferation. This is not surprising considering that altered cellular signalling and increased proliferative activity are a hall mark of the atherosclerotic disease process (Dwivedi *et al*. 2009). Further, CAD is considered as a chronic inflammatory disease where immune functions govern every stage of disease development from onset to progression (Zhou *et al*. 2005). Some of the immune-related genes have been shown to influence plaque destabilization (Breland *et al*. 2008). This could probably explain the preponderance of expression of inflammatory and immune-related genes in the present study. Similar observation has been reported in other microarray studies in CAD (Yin *et al.* 2009). Some of the differentially expressed genes identified in the current study have been previously reported in CAD such as *IL8* and *CXCR1* (Leonard *et al.* 2011), *RGS1* (Sivapalaratnam *et al.* 2012) and *PTGS2* (Cipollone and Fazia 2006).

Systems biology studies can delineate complex and dynamic molecular interactions. We generated a number of biological networks using Gene Spring software. Analysis of upregulated genes showed *IL8* to occupy a central position in the network as depicted in figure 2. A Chinese study had earlier shown the predominant involvement of inflammatory and immune-related genes through global expression profiling on peripheral lymphocytes, with marked increase in *IL8* expression (Yin *et al.* 2009). *IL8* gene has been considered as a strong chemo-attractant factor for neutrophil and T-lymphocytes (Baggiolini and Clark-Lewis 1992) and a potent promoter of angiogenesis (Li *et al.* 2003). There was strong interaction of *IL8* with *CDC42*, which is positively involved in post-infarct cardiac remodelling of mouse models (Gu *et al*. 2012), with *FOSB*, an oncogene, responsible for regulating cell proliferation, differentiation, and transformation and with *JUND*, also an oncogene which protects the cell from p53 dependent senescence and apoptosis (Weitzman *et al.* 2000).

For the purpose of independent validation by RT-PCR, we shortlisted 25 candidate genes, identified either through microarray experiment, network analysis or functional enrichment analysis. We were able to show significant differential expression of 13 out of 25 candidate genes. High differential expression was noted particularly for *IL8* and early growth response gene family (*EGR1/EGR2/EGR3*). The involvement of *IL 8* in the induction and maintenance of the inflammatory microenvironment on the vascular vessel wall is fairly well studied (Apostolakis *et al.* 2009). With regard to the early growth response gene family, the *EGR1* has been implicated in the complex modulation of vascular structure and function leading to vascular occlusive lesions (Brand *et al*. 2000). Insulin and oxidative stress has been shown to stimulate *EGR1* in vascular smooth muscle cells (Brand *et al.* 2000). On the other hand, *EGR2* plays an essential role in peripheral nerve myelination, adipogenesis and immune tolerance (Fang *et al.* 2011). Mice studies have shown the induction of *EGR3* in cardiac injury (Lyn *et al.* 2000). Also, we noted a strong correlation between *EGR3* and inflammatory genes such as *IL8*, *CXCL1*, *PTGS2* and *PTX3*, which imply that these genes may be under the influence of a common set of transcription factors and may act along shared atherothrombotic pathways. All three early growth response genes showed over 8-fold higher expression in microarray and validation studies, which underscores the importance of these genes/pathway in atherosclerosis. Further, *CDC42* also showed significant upregulation in cases in both the test datasets and validation datasets. In the present study, the pivotal role of *CDC42* was verified by network analysis where it showed complex interactions with many atherothrombotic genes. Interestingly *CCL4*, *DDX58* and *CDC42* genes showed strong correlation with each other and with other genes such as *EGR1*, that is known to regulate cell mitogenesis and differentiation (Min *et al.* 2008) and *RGS1*, which is involved in B-cell regulation and proliferation (Han *et al.* 2005). Together, these findings imply that the inflammatory genes and cell regulatory genes may actively interact and together contribute to CAD development.

A predominant portion of the differentially expressed genes in CAD as seen in our study is involved in inflammation, immune modulation, early growth response, cell signalling and proliferation. Along similar lines, Sinnaeve *et al*. (2009) reported around 160 genes to be involved in inflammation, cell growth, apoptosis from an analysis of peripheral blood gene expression in 120 cases and 121 controls. Although, the actual reported genes across both the studies are not the same, both studies demonstrate that inflammation is an important and fundamental process in CAD and therefore inflammatory genes can be a major target for treating the disease. Interestingly, Kapoor *et al*. (2014) have shown that the conventional therapies in CAD patients do not affect the expression of genes involved in inflammation and immune processes and therefore fail to provide adequate protection in CAD patients who undergo such standard therapy. Those multiple genes governing inflammatory response act in tandem to create a strong proinflammatory milieu to drive the atherosclerotic disease process was evident from the fact that inflammatory genes were upregulated in the validation samples and showed strong correlation with each other. In the present study, *IFNG* and *FASLG* genes which were downregulated in the cases also showed significant negative correlation with the inflammatory markers. While the role of *IFNG* has been associated with both proatherogenic and antiatherogenic actions (McLaren and Ramji 2009), endothelial FASLG has been shown to posses anti-inflammatory properties with inhibition of atherosclerosis under hypercholesterolaemic conditions in a study conducted on mice model, which was a cross hybridization between a *FASLG* overexpressing transgenic model and ApoE double knock out model (Yang *et al*. 2004). We also observed a positive correlation between reduced mRNA expression of *FASLG* gene and low levels of FASLG protein in plasma.

Strength and limitations of the study

One of the main shortcomings of the study has been the small sample size used for the microarray experiment. The rather high cost of consumables used in a microarray experiment

is the primary reason that precludes the use of large sample size. Therefore, the study samples were selected carefully such that the cases and controls were matched for common risk factors such as age, gender, diabetes, and hypertension, while the BMI was comparable. Age and sex are nonmodifiable risk factors for CAD. With advancing age, the chances of developing CAD increases and also is more common in men than in women. Among the modifiable risk factors, hypertension, diabetes and smoking account for 92% of the population attributable risk in the Indian population (Yusuf *et al*. 2004). In the present study, we used various permutation–combination algorithms to match these five major risk factors between cases and controls.

Another limitation is the use of blood rather than vascular tissues for gene expression analysis. Collection of vascular tissue remains a challenge whereas whole blood sample is relatively easy to obtain. Further, with increasing feasibility and reliability of high throughput genomic technologies, the need for reliable blood-based biomarkers that can report on disease states is gaining popularity. Blood contains a number of cell types such as platelets, neutrophils, lymphocytes and circulating stem cells, which directly or indirectly contribute to disease development and progression. Further, there is ample evidence to show that gene expression in peripheral blood do capture and reflect the underlying disease pathology in the vascular plaque tissue to some extent and hence can be used as surrogate tissue for gene expression studies in CAD (Smih *et al.* 2011).

Yet another limitation is the possible confounding effect of prescription drugs on gene expressions. Of all the medications routinely prescribed for CAD, statins have been known to influence gene expression, particularly lipids and inflammation (Morikawa *et al*. 2002). However, the effect of statin on inflammatory genes remains controversial (Mira and Manes 2009). Nonetheless, use of powerful bioinformatics network tools and pathway analysis of microarray findings using curated databases followed by independent validation of representative candidate genes lends strength to our study findings. Also, even with the limited sized cohort, the power of the microarray study was estimated to be around 0.9465.

In general, gene expression studies have the potential to unearth transcriptomic signatures in peripheral blood that are related to the presence and severity of CAD in the different cohorts (Kim *et al.* 2014). As a pilot study, we have looked at the differential expression between CAD cases and controls in a representative cohort of Asian Indians. Based on our early findings, we intend to undertake more focussed studies on the development of gene expression pattern in different disease states as well as relate to underlying DNA mutations that may eventually help in developing a risk prediction score (gene expression signature) for CAD.

In conclusion, in a preliminary exploration on the pattern of global gene expression, several potential candidate genes for CAD have been identified, particularly those associated with inflammation, immune modulation and early growth response. Of interest was the elevated expression of *IL8* and early growth response genes, which was identified from the microarray dataset, validated in an independent cohort and even by network analysis. Also we have reported genes like early growth response genes and *CDC42* which have not been previously reported with respect to CAD in humans. Further analysis of other candidate genes with unknown function will help unearth novel biological pathways/drug targets and provide fresh insights on molecular mechanisms governing the development and progression of CAD.

Acknowledgements

This work was supported by the Thrombosis Research Institute, London; the Department of Biotechnology, Ministry of Science and Technology, Government of India (grant number BT/01/CDE/08/07); the Tata Social Welfare Trust, India (grant number TSWT/IG/SNB/JP/Sdm); Weston foundation, UK and Foundation Bey, Switzerland. We thank all the clinical investigators, staff, administrative teams and participants of the IARS from Narayana Hrudayalaya, Bengaluru and Asian Heart Centre, Mumbai for their valuable contributions. Vinoth Kumar G., BE Computer Science (Thrombosis Research Institute, India) provided assistance in preparing the tables and figures. S. A. Deepak Ph.D. and Nilanjan Guha Ph.D. (Ms.Agilent Technologies) provided support in the experimental set up.

References

Apostolakis S., Vogiatzi K., Amanatidou V. and Spandidos D. A. 2009 Interleukin 8 and cardiovascular disease. *Cardiovasc. Res.* **84**, 353–360.

Baggiolini M. and Clark-Lewis I. 1992 Interleukin-8, a chemotactic and inflammatory cytokine. *FEBS Lett.* **307**, 97–101.

Begom R. and Singh R. B. 1995 Prevalence of coronary artery disease and its risk factors in the urban population of South and North India. *Acta Cardiol.* **50**, 227–240.

Brand E., Herrmann S. M., Nicaud V., Evans A., Ruidavets J. B., Arveiler D. *et al.* 2000 Identification of two polymorphisms in the early growth response protein-1 gene: possible association with lipid variables. *J. Mol. Med.* **78**, 81–86.

Breland U. M., Halvorsen B., Hol J., Oie E., Paulsson-Berne G., Yndestad A. *et al.* 2008 A potential role of the CXC chemokine GROalpha in atherosclerosis and plaque destabilization: down-regulatory effects of statins. *Arterioscler. Thromb. Vasc. Biol.* **28**, 1005–1011.

Cipollone F. and Fazia M. L. 2006 COX-2 and atherosclerosis. *J. Cardiovasc. Pharmacol.* **47** suppl 1, S26–S36.

Consortium I. G. V. 2005 The Indian genome variation database (IGVdb): a project overview. *Hum. Genet.* **118**, 1–11.

Dahl T. B., Yndestad A., Skjelland M., Oie E., Dahl A., Michelsen A. *et al.* 2007 Increased expression of visfatin in macrophages of human unstable carotid and coronary atherosclerosis: possible role in inflammation and plaque destabilization. *Circulation* **115**, 972–980.

Devaux Y., Bousquenaud M., Rodius S., Marie P. Y., Maskali F., Zhang L. *et al.* 2011 Transforming growth factor beta receptor 1 is a new candidate prognostic biomarker after acute myocardial infarction. *BMC Med. Genomics* **4**, 83.

Dwivedi A., Slater S. C. and George S. J. 2009 MMP-9 and -12 cause N-cadherin shedding and thereby beta-catenin signalling and vascular smooth muscle cell proliferation. *Cardiovasc. Res.* **81**, 178–186.

Edgar R., Domrachev M. and Lash A. E. 2002 Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210.

Fang F., Ooka K., Bhattacharyya S., Wei J., Wu M., Du P. *et al.* 2011 The early growth response gene Egr2 (Alias Krox20) is a novel transcriptional target of transforming growth factor-beta

that is up-regulated in systemic sclerosis and mediates profibrotic responses. *Am. J. Pathol.* **178**, 2077–2090.

Friedewald W. T., Levy R. I. and Fredrickson D. S. 1972 Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin. Chem.* **18**, 499–502.

Gu R., Zheng D., Bai J., Xie J., Dai Q. and Xu B. 2012 Altered melusin pathways involved in cardiac remodeling following acute myocardial infarction. *Cardiovasc. Pathol.* **21**, 105–111.

Han S. B., Moratz C., Huang N. N., Kelsall B., Cho H., Shi C. S. *et al.* 2005 Rgs1 and Gnai2 regulate the entrance of B lymphocytes into lymph nodes and B cell motility within lymph node follicles. *Immunity* **22**, 343–354.

Holdt L. M., Beutner F., Scholz M., Gielen S., Gabel G., Bergert H. *et al.* 2010 ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler Thromb. Vasc. Biol.* **30**, 620–627.

Huang da W., Sherman B. T. and Lempicki R. A. 2009a Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13.

Huang da W., Sherman B. T. and Lempicki R. A. 2009b Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

Kapoor D., Trikha D., Vijayvergiya R., Kaul D. and Dhawan V. 2014 Conventional therapies fail to target inflammation and immune imbalance in subjects with stable coronary artery disease: a system-based approach. *Atherosclerosis* **237**, 623–631.

Kim J., Ghasemzadeh N., Eapen D. J., Chung N. C., Storey J. D., Quyyumi A. A. *et al.* 2014 Gene expression profiles associated with acute myocardial infarction and risk of cardiovascular death. *Genome Med.* **6**, 40.

Koerselman J., Van der Graaf Y., de Jaegere P. P. and Grobbee D. E. 2003 Coronary collaterals: an important and underexposed aspect of coronary artery disease. *Circulation* **107**, 2507–2511.

Kumar N. K. 2006 Bioethics activities in India. *East. Mediterr. Health J.* **12** suppl 1, S56–S65.

Lee M. L. and Whitmore G. A. 2002 Power and sample size for DNA microarray studies. *Stat. Med.* **21**, 3543–3570.

Leonard D. A., Merhige M. E., Williams B. A. and Greene R. S. 2011 Elevated expression of the interleukin-8 receptors CXCR1 and CXCR2 in peripheral blood cells in obstructive coronary artery disease. *Coron. Artery Dis.* **22**, 491–496.

Li A., Dubey S., Varney M. L., Dave B. J. and Singh R. K. 2003 IL-8 directly enhanced endothelial cell survival, proliferation, and matrix metalloproteinases production and regulated angiogenesis. *J. Immunol.* **170**, 3369–3376.

Livak K. J. and Schmittgen T. D. 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods* **25**, 402–408.

Lyn D., Liu X., Bennett N. A. and Emmett N. L. 2000 Gene expression profile in mouse myocardium after ischemia. *Physiol. Genomics* **2**, 93–100.

Ma J. and Liew C. C. 2003 Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J. Mol. Cell. Cardiol.* **35**, 993–998.

Maitra A., Shanker J., Dash D., John S., Sannappa P. R., Rao V. S. *et al.* 2008 Polymorphisms in the IL6 gene in Asian Indian families with premature coronary artery disease–the Indian Atherosclerosis Research Study. *Thromb. Haemost.* **99**, 944–950.

Matthews L., Gopinath G., Gillespie M., Caudy M., Croft D., De Bono B. *et al.* 2009 Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622.

McLaren J. E. and Ramji D. P. 2009 Interferon gamma: a master regulator of atherosclerosis. *Cytokine Growth Factor Rev.* **20**, 125–135.

Min I. M., Pietramaggiori G., Kim F. S., Passegue E., Stevenson K. E. and Wagers A. J. 2008 The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* **2**, 380–391.

Mira E. and Manes S. 2009 Immunomodulatory and anti-inflammatory activities of statins. *Endocr. Metab. Immune Disord. Drug Targets* **9**, 237–247.

Morikawa S., Takabe W., Mataki C., Kanke T., Itoh T., Wada Y. *et al.* 2002 The effect of statins on mRNA levels of genes related to inflammation, coagulation, and vascular constriction in HUVEC. Human umbilical vein endothelial cells. *J. Atheroscler. Thromb.* **9**, 178–183.

Shanker J., Maitra A., Rao V. S., Mundkur L., Dhanalakshmi B., Hebbagodi S. *et al.* 2010 Rationale, design & preliminary findings of the Indian Atherosclerosis Research Study. *Indian Heart J.* **62**, 286–295.

Sinnaeve P. R., Donahue M. P., Grass P., Seo D., Vonderscher J., Chibout S. D. *et al.* 2009 Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. *PLoS One* **4**, e7037.

Sivapalaratnam S., Basart H., Watkins N. A., Maiwald S., Rendon A., Krishnan U. *et al.* 2012 Monocyte gene expression signature of patients with early onset coronary artery disease. *PLoS One* **7**, e32166.

Smih F., Desmoulin F., Berry M., Turkieh A., Harmancey R., Iacovoni J. *et al.* 2011 Blood signature of pre-heart failure: a microarrays study. *PLoS One* **6**, e20414.

Tillin T., Dhutia H., Chambers J., Malik I., Coady E., Mayet J. *et al.* 2008 South Asian men have different patterns of coronary artery disease when compared with European men. *Int. J. Cardiol.* **129**, 406–413.

Vinukonda G., Shaik Mohammad N., Md Nurul Jain J., Prasad Chintakindi K. and Rama Devi Akella R. 2009 Genetic and environmental influences on total plasma homocysteine and coronary artery disease (CAD) risk among South Indians. *Clin. Chim. Acta* **405**, 127–131.

Weitzman J. B., Fiette L., Matsuo K. and Yaniv M. 2000 JunD protects cells from p53-dependent senescence and apoptosis. *Mol. Cell.* **6**, 1109–1119.

Yang J., Sato K., Aprahamian T., Brown N. J., Hutcheson J., Bialik A. *et al.* 2004 Endothelial overexpression of Fas ligand decreases atherosclerosis in apolipoprotein E-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1466–1473.

Yin H. M. X., Jiang Y, Shi D and Chen K. 2009 Investigation of gene expression profiles in coronary heart disease and functional analysis of target gene. *Chin. Sci. Bull.* **54**, 759–765.

Yusuf S., Hawken S., Ounpuu S., Dans T., Avezum A., Lanas F. *et al.* 2004 Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* **364**, 937–952.

Zhou X., Robertson A. K., Rudling M., Parini P. and Hansson G. K. 2005 Lesion development and response to immunization reveal a complex role for CD4 in atherosclerosis. *Circ. Res.* **96**, 427–434.

Received 2 January 2015, in revised form 27 February 2015; accepted 13 March 2015
Unedited version published online: 16 March 2015
Final version published online: 13 October 2015