Determinants of Excess Genetic Risk of Acute Myocardial Infarction – A Matched Case-control Study

Zdeněk Valenta[1], **Ivan Mazura**[1], **Michal Kolář**[1], **Hana Grünfeldová**[2], **Petra Feglarová**[1], **Jan Peleška**[1], **Marie Tomečková**[1], **Jan Kalina**[1], **Dalibor Slovák**[1], **Jana Zvárová**[1]

[1] Center of Biomedical Informatics, Institute of Computer Science AS CR, Prague, Czech Republic

[2] Centre of Biomedical Informatics, Municipal Hospital Čáslav, Czech Republic

Summary

Background: Myocardial infarction and stroke represent a major public health problem in most developing countries. This study explores genetic predisposition of acute myocardial infarction in the Czech population.
Methods and Results: Genome-wide expression study used matched case-control design. Peripheral blood samples of the controls were matched to those of cases based on gender, age, status of diabetes mellitus and smoking status. Six months cardiovascular survival status of the cases was used to identify two distinct subgroups among the cases. Linear models for microarray data were employed to identify differential gene expression. Shrunken centroids technique helped in identifying the subsets of differentially expressed genes with predictive properties in independent samples. Predictive properties were evaluated using bootstrap sampling. Sixty transcripts were found to be both clinically and statistically differentially expressed among the cases not surviving the six months follow-up period relative to controls, while no such transcripts were observed among other surviving cases. The two subgroups of cases exhibited fourteen differentially expressed transcripts. Predictive modeling indicated sixteen out of sixty transcripts to best discriminate between the controls and cases that died during the follow-up period from cardiovascular causes, while for the surviving cases the already non-significant set of transcripts could not be further reduced. Eleven out of fourteen transcripts were found to best discriminate between the two groups of cases using shrunken centroids.
Conclusions: The study identified genes associated with excess genetic risk of acute myocardial infarction, including those associated with the six months fatality of the cases.

Keywords

genome-wide association study, gene expression; myocardial infarction, genetic predisposition, predictive modeling

Correspondence to:

Zdeněk Valenta
Dept. of Medical Informatics and Biostatistics
Institute of Computer Science AS CR
Address: Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic
E–mail: valenta@cs.cas.cz

EJBI 2012; 8(1):34–43
received: October 22, 2012
accepted: November 25, 2012
published: December 30, 2012

1 Introduction

Morbidity and mortality from atherosclerotic complications, such as acute myocardial infarction and stroke, continue to represent major public health issue in most developing countries. They are caused by multiple environmental and genetic factors and the interaction between them. While there are clinical risk factors known to be associated with the incidence of acute myocardial infarction, genetic profile of an individual may represent additional factors independently associated with the incidence of this outcome.

Genome-wide expression profiling provides comprehensive summary of mRNA levels in a tissue sample, allowing for identification of the sets of genes and transcripts associated with individual condition. Microarray studies of human diseases are often limited by challenges in obtaining human tissues. Peripheral blood has become an attractive prime tissue for biomarker detection because of its critical role in immune response, metabolism, communication with cells and the extracellular matrix in almost all tissues and organs in the human body, as well as for the simplicity and low invasiveness of sample collection [1].

2 Study Design and Methods

Experimental design of this study aimed at identifying the genes associated with excess genetic risk for the incidence of acute myocardial infarction which is not necessarily captured through known clinical risk factors.

Forty five cases with confirmed diagnosis of acute myocardial infarction were enrolled between September 2006 and January 2011. The diagnosis had to satisfy the clinical criteria, ECG outcome and laboratory findings according to medical guidelines. Coronary angiography was performed in most patients. The cases had to be less than 80 years old and no subjects could be actively treated for cancer.

Venous blood samples were drawn from each subject enrolled. Paired controls were selected out of patients hospitalized for motoric complications with no evidence of coronary artery or peripheral artery disease, normal ECG and no history of stroke. They were matched to cases based on their gender, age (the controls could be up to 5 years older than cases), status of Type II diabetes mellitus and smoking status. These variables represent clinical and underlying genetic factors associated with the incidence of acute myocardial infarction. This study focuses on identifying the profiles associated with excess genetic risk which are not necessarily expressed through these risk factors.

Six months following the cardiac event cardiovascular survival status was assessed for all the cases. We hypothesized that the cases who have not survived the six months follow-up period (AMID6) and those who did (AMI) would each differ in their genetic make-up from the controls. We also hypothesized differences in genetic profiles between the two groups of cases (AIMD6 vs. AMI). While the average paired gene expression differences between the cases and their corresponding matched controls capture the primary prevention perspective, the differences between AMID6 and AMI reflect the secondary prevention point of view.

The study complies with the Declaration of Helsinki and was approved by the local ethics committee. All participants gave written informed consent. Basic descriptive characteristics of the data are provided in Tables 1 and 2 below and supplementary Table S10 in [44].

3 Microarray Analysis

The study utilized Illumina microarray technology for analyzing gene expression intensities across the full human genome. Samples of peripheral whole blood were collected from all subjects using commercial 3 ml Vacutainer® sets with EDTA. The tubes were inverted several times and 2.4 ml of the content was immediately mixed with 7.6 ml RNAlater® (Applied Biosystems) in 15 ml tubes, stabilized blood samples were inverted several times until they were homogenous and the samples were stored in -70°C. The RNA was isolated from 1.8 ml aliquot of stabilized blood with RiboPure^{TM}-Blood Kit (Applied Biosystems, U.S.A.), precipitated and purified with GLOBINclear^{TM}–Human kit (Applied Biosystems, U.S.A.). The quantification was made on Nanodrop (Thermo Scientific, U.S.A.) and the integrity of the RNA was measured on Bioanalyzer 2100 (Agilent Technologies Inc., U.S.A.). The cRNA was amplified using Illumina® TotalPrep^{TM} RNA Amplification Kit, precipitated and controlled on Nanodrop (Thermo Scientific, U.S.A.) and Bioanalyzer 2100 (Agilent Technologies Inc., U.S.A.). The cRNA samples (1.5 μg) were hybridized on HumanWG-6 v2 Expression BeadChips (Illumina Inc., U.S.A.) according to manufacturer's protocol.

4 Statistical Analysis

Statistical analysis used the R system for statistical computing, graphics and data analysis [2]. We used several packages which are part of the Bioconductor project [3]. The 'beadarray' package [4] was used for reading in the gene expression data from Illumina analyzer scans, the 'BASH' method [5] was used to identify defective beads on Illumina chips. We adopted 'normal-exponential convolution' method [6] for separating background noise from the signal. The $\log_2$-transformed quantile-normalized gene expression intensities were modeled using two explanatory variables, the matched pair indicator and the sampling group indicator ('AMI', 'AMID6', 'Controls') using the 'limma' package accounting for correlated data due to several biologically replicated samples [7]. Applying the empirical Bayes approach to model fit rendered moderated t-tests for each transcript/gene and contrast of interest. Multiple testing issues were handled using the q-value approach [8]. The two principal contrasts of interest estimated the mean paired differences in gene expression intensities between the cases (AIMD6, AIM) and their matched controls, respectively. Of interest were also the gene expression differences between the two groups of cases. Statistical significance was reached for transcripts with q-value below 0.05, clinical significance was reached when the $\log_2$-fold change was greater or equal to 1 in absolute value.

To identify subsets of genes possessing predictive properties in independent samples we employed shrunken centroids approach [9] implemented in the 'Predictive Analysis for Microarrays' (PAM). Subsets of genes identified as differentially expressed using the limma package were further analyzed using PAM. The final sets of genes so identified are believed to possess predictive properties in independent samples, which were evaluated using bootstrapping. PAM modeling technique was also applied to the full genome.

5 RT-qPCR Validation

Modeling results were validated by RT-qPCR analysis which used available RNA samples from the four cases, their matched controls and 6 other randomly selected controls. The genes ADORA3, VNN3, IL18R1, IL18RAP, ERLIN1, FOS and ARG1 were quantified while 18S and HPRT were used as housekeeping references for each tested sample. Gene SPATC1 was selected as negative control.

6 Results

6.1 Limma Modeling Results

Comparing the matched controls with the cases who died from cardiovascular causes within six months following the cardiac event (contrast 'AMID6 vs. Controls') implicated 60 differentially expressed genes/transcripts which met the criteria of both statistical ($q < 0.05$) and clinical ($|log_2FC| \geq 1$) significance. Of those genes, 40 were up- and 20 down-regulated. Without regard to clinical significance, statistical significance was attained for 323 transcripts; see the Venn diagrams in Figure 1.

Comparing the cases who survived the 6 months follow up period with their matched pair controls (contrast 'AMI vs. Controls') revealed no genes that would meet the above mentioned criteria for either statistically or clinically significant differential expression.

The population gene expression differences between the two groups of cases (contrast 'AMID6 vs. AMI') were associated with 14 transcripts which met the criteria of both statistical and clinical significance, out of which 4 were up- and 10 down-regulated. Statistical significance was observed for 60 genes/transcripts, out of which 13 were up- and 47 down-regulated.

Supplementary tables S1, S4 and S7 in [44] present a detailed view of limma modeling results assessing the three linear contrasts in gene expression intensities.

6.2 Predictive PAM Modeling

For all three types of contrasts considered in this study we were particularly looking for the candidate sets of genes which would possess predictive properties in independent samples. Therefore, the available samples were studied further using shrunken centroids approach. The corresponding results are summarized in Table 3. With each

Table 1: Group counts and percentages for categorical variables.

Variable	Level	Group Counts and Percentages		
		AMI	AMID6	CONTROL
Gender	Male	28 (68%)	2 (50%)	30 (67%)
Smoking	Smokers	10 (24%)	0 (0%)	10 (22%)
Type 2 DM	YES	12 (29%)	2 (50%)	14 (31%)
Dyslipidemia	YES	17 (41%)	2 (50%)	15 (33%)
Hypertension	YES	32 (78%)	2 (50%)	29 (64%)
First AMI	YES	32 (78%)	4 (100%)	–
STEMI	YES	26 (63%)	3 (75%)	–
Heart failure	YES	5 (12%)	3 (75%)	–
PCI	YES	6 (15%)	0 (0%)	–
ACEI*	YES	21 (51%)	1 (25%)	16 (36%)
Betablockers*	YES	19 (46%)	0 (0%)	15 (33%)
Diuretics*	YES	11 (27%)	1 (25%)	14 (31%)
Ca blockers*	YES	12 (29%)	3 (75%)	11 (24%)
Statins*	YES	14 (34%)	2 (50%)	15 (33%)
Fibrates*	YES	2 (5%)	0 (0%)	2 (4%)
Other*	YES	20 (49%)	3 (75%)	36 (80%)

*) Chronic medication

Table 2: Descriptive characteristics of continuous variables.

Variable	Group Means and Standard Deviations*		
	AMI	AMID6	CONTROL
Age	63.6 (9.18)	72.3 (4.73)	65.5 (9.42)
Height (cm)	167.3 (10.0)	163.3 (10.4)	165.4 (10.2)
Weight (kg)	85.2 (17.9)	81.0 (17.0)	80.3 (12.1)
SBP (mmHg)	140.0 (27.8)	137.5 (12.6)	142.3 (18.0)
DBP (mmHg)	82.7 (15.8)	84.3 (16.5)	82.8 (9.2)

*) At ICU entry

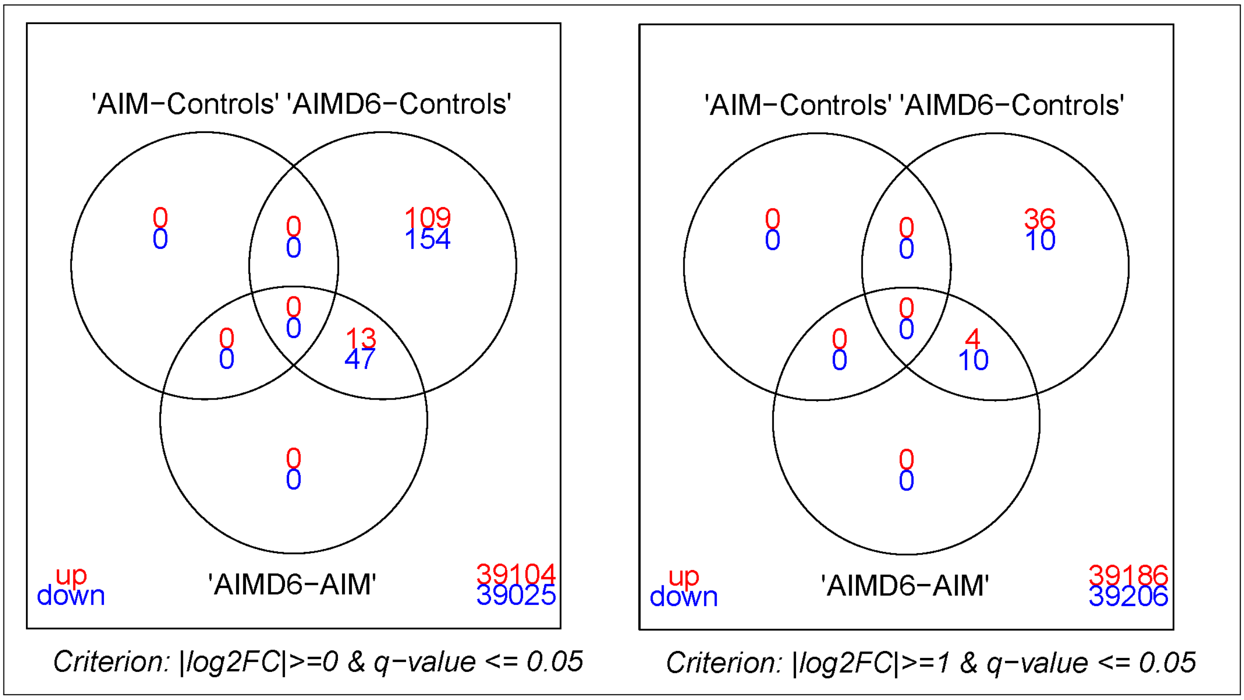


Figure 1: Counts of statistically significant differentially expressed transcripts (left panel) and those reflecting both statistical and clinical significance (right panel).

contrast Table 3 presents two candidate sets of genes. The Set ♯1 resulted from further reducing the gene set obtained from the limma modeling using PAM while the Set ♯2 was obtained by applying the PAM technique to all 39 226 available transcripts. Grey shading accentuates simultaneous presence of both statistical and clinical significance, as defined for this study. Comprehensive results covering the full genome reduction using PAM may be found in supplementary tables S3, S6 and S9 in [44], tables S2, S5 and S8 in [44] show PAM reduction of the gene sets obtained using limma.

6.3 AMID6 vs. Controls

Applying the shrunken centroids technique to the set of both statistically and clinically significant transcripts identified via limma modeling (top 60 genes of Table S1 in [44]) rendered 16 genes (Set ♯1) with predictive properties in independent samples shown in Table 3. Initiating the PAM modeling with the full genome resulted in a set of 14 genes (Set ♯2) of which only IL18R1 and DUSP1 would not pass the criterion of statistical significance within limma modeling framework while adhering to the clinical one in all instances. Remarkably, the two corresponding sets have a large proportion of genes in common.

6.4 AMI vs. Controls

The set of genes obtained using limma modeling for this contrast (Table S4 in [44]) did not exhibit statistical or clinical significance as defined for our study and could not be further reduced using PAM modeling. The set obtained by applying the PAM technique to the full genome was quite extensive, counting 228 transcripts. Table 3 presents truncated list of top thirteen genes (Set ♯2) which includes five genes from the corresponding Set ♯1.

6.5 AMID6 vs. AMI

PAM reduction of 14 statistically and clinically significant transcripts identified via limma modeling (see Table S7 in [44]) rendered 11 predictive transcripts, two of which pertain to gene 'CLYBL'. Initiating the PAM analysis with the full genome resulted in a set of 22 transcripts, four of which also appeared in the corresponding Set ♯1.

Table 4 presents estimates of sensitivity and specificity of the PAM classifier obtained from three bootstrap studies evaluating predictive properties of the two sets of genes identified for each contrast of interest. The studies used 1000 samples from the target population with replacement. We report the mean values along with the 5^{th} and 95^{th} percentile for both quantities of interest.

Predictive properties assessed using the PAM classifier appeared generally more favorable when the gene sets ♯2 were employed while notable improvements were observed in relation to sensitivity rather than specificity. Some improvements, however, came at a price of large number of genes required. This was the case of contrast 'AMI vs. Controls' where the Set ♯1 was of size 13 while the Set ♯2 counted 228 transcripts. This is a likely consequence of having observed no statistically or clinically significant

Table 3: Predictive sets of genes identified using PAM for each contrast of interest, with the ranks in the Set ♯1 and Set ♯2, respectively, q-values and $\log_2$FC based on limma results.

Symbol	**Ref Seq ID**	**Definition**	**Set ♯1 Rank***	**Set ♯2 Rank†**	**q-value**	**$\log_2$FC**
Contrast AMID6 vs Controls						
ECHDC3	NM$_0$24693.2	enoyl Coenzyme A hydratase domain containing 3	1	1	0.0498	2.03
IL18RAP	NM_003853.2	interleukin 18 receptor accessory protein	2	3	0.0195	1.28
PFKFB2	NM_006212.2	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	3	5	0.0085	1.93
IRS2	NM_003749.2	insulin receptor substrate 2	4	6	0.0308	1.43
PHACTR1	NM_030948.1	phosphatase and actin regulator 1	5	4	0.0352	1.86
ERLIN1	NM_006459.2	ER lipid raft associated 1	6	2	0.0416	1.77
VNN3	NM_001024460.1	vanin 3	7	7	0.0290	1.44
ADORA3	NM_020683.5	adenosine A3 receptor	8	9	0.0525	2.10
CLEC4E	NM_014358.1	C-type lectin domain family 4, member E	9	11	0.0288	1.75
ASPRV1	NM_152792.1	aspartic peptidase, retroviral-like 1	10	12	0.0352	1.01
PFKFB2	NM_001018053.1	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	11	5	0.0416	1.69
CPD	NM_001304.3	carboxypeptidase D	12	10	0.0397	1.31
FKBP5	NM_004117.2	FK506 binding protein 5	13	—	0.0596	1.03
PRKDC	NM_006904.6	protein kinase, DNA-activated, catalytic polypeptide	14	—	0.0075	1.10
NPM1	NM_199185.1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	15	—	0.0223	–1.15
SAMSN1	NM_022136.3	SAM domain, SH3 domain and nuclear localization signals 1	16	—	0.0478	1.37
IL18R1	NM_003855.2	interleukin 18 receptor 1	—	8	0.0525	2.10
DUSP1	NM_004417.2	dual specificity phosphatase 1	—	13	0.0596	1.03
LOC645649	XM_928663.1	hypothetical protein LOC645649	—	14	0.0075	1.10
Contrast AMI vs Controls						
OLIG2	NM_005806.2	oligodendrocyte lineage transcription factor 2	1	1	0.0756	–0.89
VNN3	NM_001024460.1	vanin 3, transcript variant 3	2	2	0.0756	0.50
MS4A3	NM_006138.4	membrane-spanning 4-domains, subfamily A, member 3, transcript variant 1	3	6,7	0.0756	–0.64
CEBPE	NM_001805.2	CCAAT/enhancer binding protein (C/EBP), epsilon	4	5	0.0756	–0.45
FOS	NM_005252.2	v-fos FBJ murine osteosarcoma viral oncogene homolog	5	4	0.0756	0.39
LIPA	NM_000235.2	lipase A, lysosomal acid, cholesterol esterase, transcript variant 2	6	—	0.0756	–0.37
LOC645649	XM_928663.1	hypothetical protein LOC645649	7	—	0.0906	0.29
TCRB	M97723	T cell receptor beta locus	8	—	0.0756	0.38
EPAS1	NM_001430.3	endothelial PAS domain protein 1	9	—	0.0756	–0.31
CLINT1	NM_014666.2	clathrin interactor 1	10	—	0.0756	–0.25
MYCT1	NM_025107.1	myc target 1	11	—	0.0756	–0.15
VPS29	NM_016226.2	vacuolar protein sorting 29 (yeast), transcript variant 1	12	—	0.0756	–0.15
LOC130951	NM_138804.2	hypothetical protein BC014602	13	—	0.0756	–0.13
CCL23	NM_005064.3	chemokine (C-C motif) ligand 23, transcript variant CKbeta8-1	—	3	0.2633	–0.84
MYB	NM_005375.2	v-myb myeloblastosis viral oncogene homolog (avian)	—	8	0.1397	–0.47
C13orf18	NM_025113.1	chromosome 13 open reading frame 18	—	9	0.1347	0.40
PER1	NM_002616.1	period homolog 1 (Drosophila)	—	10	0.2012	0.31
CCL23	NM_005064.3	chemokine (C-C motif) ligand 23, transcript variant CKbeta8-1	—	11	0.3385	–0.50
OLIG1	NM_138983.1	oligodendrocyte transcription factor 1	—	12	0.1347	–0.49
PRSS33(**)	NM_152891.1	protease, serine, 33	—	13	0.3606	–0.58
Contrast AMID6 vs AMI						
ADORA3	NM_020683.5	adenosine A3 receptor	1	15	0.0490	1.78
TCRB	M97723	T cell receptor beta locus	2	8	0.0130	–1.52
ERLIN1	NM_006459.2	ER lipid raft associated 1	3	11	0.0490	1.19
CLYBL	NM_206808.1	citrate lyase beta like	4	19	0.0234	–1.08
TCEA3	NM_003196.1	transcription elongation factor A (SII), 3	5	—	0.0381	-1.66
TCRA	BC070337	T cell receptor alpha locus	6	—	0.049	–1.42
CLYBL	NM_206808.1	citrate lyase beta like	7	19	0.0130	–1.18
HSD17B8	NM_014234.3	hydroxysteroid (17-beta) dehydrogenase 8	8	—	0.0490	–1.06
FLT3	NM_004119.1	fms-related tyrosine kinase 3	9	—	0.0490	1.14
AXIN2	NM_004655.2	axin 2 (conductin, axil)	10	—	0.0388	–1.49
—	CR596519	full-length cDNA clone CS0DI056YK21 of Placenta Cot 25-normalized	11	—	0.0490	–1.45
BCAT1	NM_005504.4	branched chain	—	1	0.1146	1.05
—	AW337887	he12d07.x1 NCI_CGAP_CML1 cDNA clone IMAGE:2918797 3’	—	2	0.0130	0.84
AMPH	NM_001635.2	amphiphysin (AMPH), transcript variant 1	—	3	0.0814	1.20
—	BM682470	UI-E-EJ0-aig-b-14-0-UI.s1 UI-E-EJ0, cDNA clone UI-E-EJ0-aig-b-14-0-UI 3’	—	4	0.0490	–0.76
C7orf53	NM_182597.1	chromosome 7 open reading frame 53	—	5	0.1371	0.86
—	CR592039	full-length cDNA clone CS0CAP005YH21 of Thymus	—	6	0.0994	–1.48
C2orf58	NM_173652.1	chromosome 2 open reading frame 58	—	7	0.1053	0.78
ASPRV1	NM_152792.1	aspartic peptidase, retroviral-like 1	—	9	0.0847	1.19
—	CN484989	hx21e11.y1 primary human ocular pericytes. Equalized (hx) cDNA clone hx21e11 5’	—	10	0.0721	1.44
ETS2	NM_005239.4	v-ets erythroblastosis virus E26 oncogene homolog 2	—	12	0.1218	0.73
NDUFB3	NM_002491.1	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	—	13	0.4508	0.43
—	X00437	mRNA for T-cell specific protein	—	14	0.0721	–1.44
ZNF516	XM_496278.2	zinc finger protein 516	—	16	0.2826	0.48
SLC26A8	NM_052961.2	solute carrier family 26, member 8, transcript variant 1	—	17	0.0721	0.63
KIF20B	NM_016195.2	kinesin family member 20B	—	18	0.0130	–0.86
ARG1	NM_000045.2	arginase, liver	—	20	0.0814	3.05
IL18RAP	NM_003853.2	interleukin 18 receptor accessory protein	—	21	0.0542	1.61
CD59	NM_203329.1	CD59 molecule, complement regulatory protein, transcript variant 3	—	22	0.2374	0.65

*) PAM reduction of the set of statistically and clinically significant genes obtained from limma modeling

†) PAM reduction of the whole genome (represented by 39,226 transcripts)

**) Truncated list of genes

transcripts using the limma modeling (see Table S4 in [44]).

6.6 RT-qPCR Validation Study

Results of the RT-qPCR validation are summarized in supplementary Table S12 in [44]. One observation in the AIMD6 group of cases (ID=C078) appeared fairly similar to controls, its removal lead to the overall improvement. Details concerning the subjects from AMID6 group are shown in supplementary Table S11 and S12 in [44] presents the summary of validation study results.

7 Discussion

7.1 AMI vs. Controls

This contrast identifies genes which were differentially expressed between general population and the cases who survived the six months follow-up period following the MI episode. Although no statistically significant results were obtained for this contrast, several genes shown in Table 3 were previously linked to MI and cardiovascular or coronary artery disease. The FOS gene plays a role in functional organization of central cardiovascular pathways; its expression in certain central neurons may lead to sustained changes in cardiovascular function [10]. VNN3 is part of pantetheinase gene family which regulates hydrolysis of pantetheine into pantothenic acid (vitamin B5) and cysteamine, a potent antioxidant. Human neutrophils express transcripts encoding multiple splice variants of VNN3 [11]. In relation to oxidative stress VNN3 may play a role in tissue repair [12]. Phosphorylation of clock protein PER1 was shown to regulate its circadian degradation in normal human fibroblasts [13]. Mutations in the LIPA gene were shown to be related to cholesterol metabolism [14] and the gene was described as susceptibility gene for the incidence of coronary artery disease [15, 16]. Genes OLIG1 and OLIG2 encode transcription factors expressed in both the developing and mature vertebrate central nervous system and may have additional functions in a variety of neurological diseases [17]. Wojakowski et al [18] report up-regulation of OLIG2 in stroke patients. Inouye et al [19] report gene MS4A3 as one of only three strong predictors of lipid leukocyte module from their genome-wide study. Other genes showing significant evidence of association with lipid traits identified also in our study using PAM reduction of the whole genome include GATA2, CPA3, C1ORF186, C1ORF150, SLC45A3, SPRYD5 and CEBPD (Supplementary Table S6 in [44]), all with potential contribution to the pathogenesis of coronary artery disease. Gene EPAS1 was identified as a significant promoter of angiogenesis [20]. Castillo et al [21] demonstrated that inflammatory chemokine CCL23 is independently associated with coronary atherosclerosis. The MYB gene plays essential role in adult vascular smooth muscle cells survival [22].

7.2 AMID6 vs. Controls

This contrast targets differentially expressed genes among the cases not surviving the 6 months follow-up relative to general population. All predictive genes of the Set ♯2 in Table 3 have shown both clinically and statistically significant differential expression based on limma modeling. Gene ADORA3 is known as a receptor mediating cardioprotective functions during ischemia [23]. In our study the gene was overexpressed in cases that died within 6 months from cardiovascular complications relative to controls. The same was true relative to surviving cases ('AIMD6 vs. AMI') while the gene was under expressed among surviving cases relative to controls (supplementary Table S6 in [44]). Increased activity may be observed with genes involved in the overall immune response (IL18R1, IL18RAP). Liangos et al [24] uncovered highly intertwined signaling underlying ischemia reperfusion and inflammatory response. The corresponding genes identified in our study include IL18R1, IL18RAP, IL1RAP, LCN2 and TLR4 (tables S1 and S6 in [44]). Mallat et al [25] report significant expression of proinflammatory cytokine IL-18 and its signaling receptor IL-18R in human atherosclerotic plaques. Rosenberg et al [26] validated diagnostic test based on the expression of 23 genes previously found to be associated with the presence of CAD. From the genes they used in predictive modeling those identified in our study involve IL18RAP, TLR4 and CLEC4E, a mediator of immune and inflammatory re-

Table 4: Bootstrap estimates of sensitivity and specificity of the PAM classifier.

		Predictive Set ♯1			**Predictive Set ♯2**		
		(based on limma results)			(based on 39 226 transcripts)		
Contrast	**Item**	**Mean**	**5%**	**95%**	**Mean**	**5%**	**95%**
AMID6 vs. Controls	Sensitivity	0.90	0.75	1.00	1.00	1.00	1.00
	Specificity	0.93	0.84	1.00	0.96	0.87	1.00
AMI vs. Controls	Sensitivity	0.73	0.63	0.83	0.89	0.78	0.98
	Specificity	0.87	0.80	0.93	0.85	0.76	0.96
AMID6 vs. AMI	Sensitivity	0.89	0.50	1.00	1.00	1.00	1.00
	Specificity	0.95	0.90	1.00	0.96	0.90	1.00

sponse. Tiret et al [27] related genetic variability in IL18, IL18R1 and IL18RAP to cardiovascular mortality. The gene SAMSN1 was found to be preferentially expressed in mast cells [28] containing large amounts of heparin and histamine. Protein encoded by the PFKFB2 gene mediates control of glycolysis in eukaryotes. IRS2 gene was shown to be associated with severe obesity and insulin sensitivity in Type II diabetic patients [29, 30]. The gene VNN3 reappears also with this contrast. Gene PHACTR1 was cited for the association with CAD, CVD and MI based on several recent genome-wide studies [31, 32]. The ERLIN1 gene was recently identified as a member of the prohibitin family of proteins that define lipid-raft-like domains of the ER [33]. Polymorphisms in FKBP5 may be associated with increased vulnerability to posttraumatic stress disorder [34]. Gene PRKDC is a central regulator of DNA double-strand break repair. Down-regulation of NPM1 was previously linked to cardiac cell differentiation [35], the DUSP1 gene was found to be associated with oxidative stress response [36]. No references in relation to CVD, CAD or MI could be found for the genes ECHDC3 and ASPRV1.

7.3 AMID6 vs. AMI

This contrast signifies population gene expression differences between the surviving cases and those who died within 6 months following the acute MI episode. Both statistical and clinical significance based on limma modeling was attained for Set ♯1. Several predictive genes are repeated here from the 'AIMD6 vs. Controls' contrast. They include ADORA3, IL18RAP, ERLIN1, ASPRV1, gene TCRB is repeated from 'AIM vs. Controls'. Strongly down-regulated genes TCRA, TCRB and AXIN2 participate in V(D)J recombination, T-cell and leukocyte differentiation, antibody-dependent cellular cytotoxicity and signal transduction. Dumont et al [37] report association of ARG1 polymorphisms with the risk of AMI and common carotid intima media thickness. Harpster et al [38] report ARG1 as the single most highly induced transcript in post-myocardial infarction subjects. Complement regulator CD59 is a potent inhibitor of the membrane attack complex (MAC). Acosta et al suggest that in diabetes glycation–inactivation of endothelial CD59 would contribute to the development of vascular complications [39]. CD59 was shown to protect against atherosclerosis by restricting the MAC formation [40]. Transcription factor ETS2 was recently identified to determine inflammatory state of endothelial cells in advanced atherosclerotic lesions [41]. Meta-analysis of 15 GWAS studies [42] revealed a few genes associated with resting heart rate, a predictor of cardiovascular mortality, including BCAT1 gene.

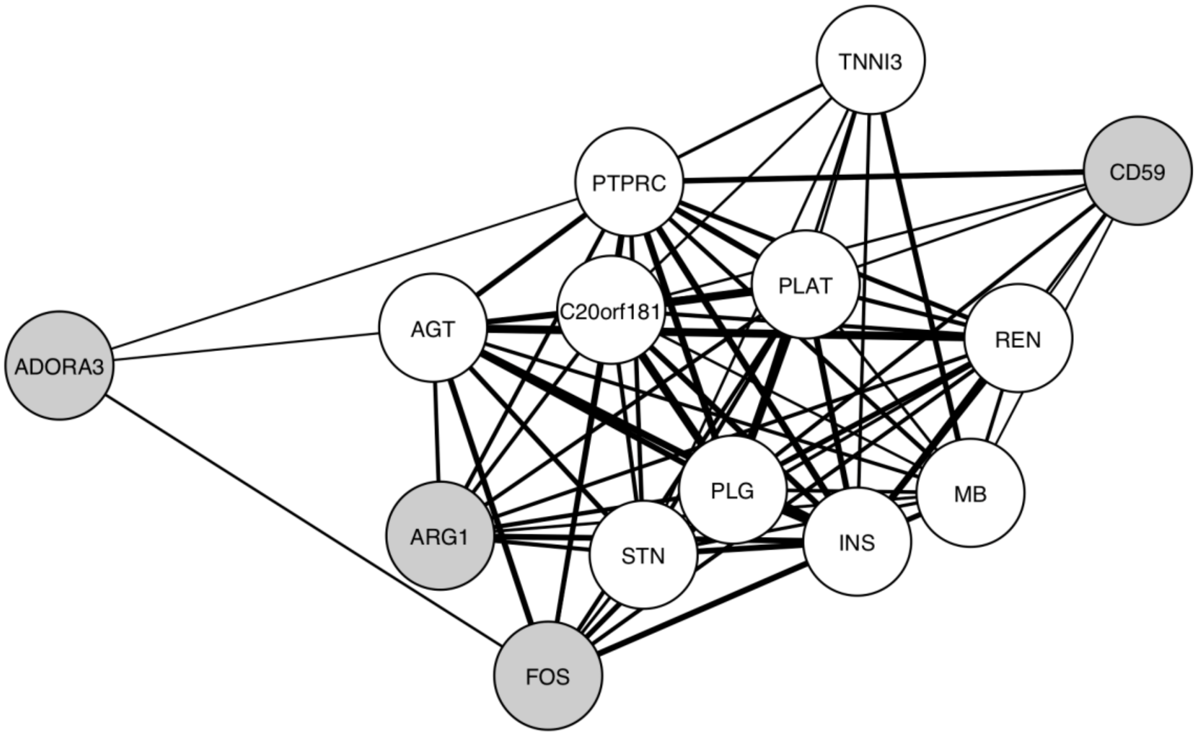
Number of genes and transcripts identified in our study as being associated with the outcome represent novel candidates which were not previously linked to the incidence of acute myocardial infarction. We illustrate four up-regulated (AMPH, FLT3, ZNF516) and five down-regulated genes (AXIN2, CLYBL, KIF20B, TCEA3, TCRA) identified in our study. Amphiphysin-synaptic vesicle-associated protein (AMPH) observed in Stiff-Man syndrome includes SH3 domains in C-terminal region. Up-regulated activity of the gene FLT3 is linked to hematopoiesis activation, angiogenesis, hematopoietic progenitor cell differentiation, macrophage differentiation and interleukin, natural killer activation. ZNF516 (zinc finger protein 516) is a gene family member, coordinating Zn-ions in stabilizing different cellular processes. AXIN2 plays important role in beta-catenin stabilization. CLYBL encodes beta-like citrate lyase. KIF20B, kinesin family member 20B, is a structure required for completion of cytokinesis. The group of down-regulated genes includes the gene TCEA3 providing interaction with the enzyme RNA polymerase II during the transcription process. TCRA is a T-cell antigen receptor, alpha-subunit.

Furthermore, we identified novel up-regulated (LOC645649, c13orf18, AW337887, c7orf53, c2orf58, CN484989) and down-regulated structures (LOC130951, CR596519, BM682470, CR592039, X00437) that are listed in Table 3. Recently, Puigdecanet et al [43] identified C13orf18 being part of a molecular signature characterized with an upregulation of inflammatory genes related to neutrophil activation and thrombosis.

Text mining search of medical literature performed at PubGene.com using MeSH term 'Myocardial Infarction' and the set of genes found to be predictive for MI (see Table 3) rendered 10 genes most cited for their association with MI plus four genes (ADORA3, FOS, ARG1, CD 59) indicated in our study which appear to be related to these genes. Figure 2 shows that the four genes are linked to positive regulation of Interleukin 12 production (ADORA3), co-regulation of insulin secretion (FOS), regulation of reverse cholesterol transport, co-regulation of insulin secretion, cholesterol absorption, cardiac muscle contraction and glycoprotein biosynthetic process (ARG1) and activation of membrane attack complex (CD59). Supplementary Figure S1 in [44] uncovers relationships of these four genes with other genes. Biological processes associated with the genes shown in Table 3 are summarized in supplementary Table S13 in [44].

Only a partial agreement with the gene sets reported to be associated with the incidence of MI from other studies may in part be explained by the differences in statistical design, studied population and respective sample sizes, use of non-homogenous subgroups (diabetics, nondiabetic patients), population distributions of related risk factors, therapy (especially use of statins), exclusion criteria, existence of concomitant diseases and other inflammatory conditions, heart failure, smoking, extent of non-coronary atherosclerosis, and examination of circulating cells and other cells in tissues.

Due to significant costs involved in microarray analysis our study is limited by a relatively small sample size. Synthesis of the genetic and clinical information gathered from genomic studies is expected to refine personalized approaches to managing the risk of CAD. Genetic risk scores



Note: The thickness of the lines connecting individual genes corresponds to the number of articles on the $\log_{10}$ scale citing the association.
Data source http://www.pubgene.com.

Figure 2: Literature search of 10 genes most frequently cited for their association with MI and their relation to predictive genes identified in Table 3.

derived from several functional single nucleotide polymorphisms (SNPs) or haplotypes in multiple genes may improve the prediction of CAD.

Acknowledgments

All persons who have made substantial contributions in the manuscript have fulfilled the authorship criteria and are listed as (co)-authors of the paper.

Funding Sources

The work on this project was supported by the grant 1M06014 from the Ministry of Education, Youth and Sports of the Czech Republic.

Disclosures

The authors declare no conflict of interest.

References

[1] Taurino C. Miller WH, McBride MW, McClure JD, Khanin R, Moreno MU, Dymott JA, Delles C, Dominiczak AF. Gene expression profiling in whole blood of patients with coronary artery disease. Clin Sci (Lond). 2010 Jul 9;119(8):335-43

[2] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2010, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[3] Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology. 2004; 5(10):R80.

[4] Dunning MJ, Smith ML, Ritchie ME, Tavaré S. beadarray: R Classes and Methods for Illumina Bead-based Data. Bioinformatics. 2007; 23(16):2183–4, 2007.

[5] Cairns JM, Dunning MJ, Ritchie ME, Russell R., Lynch AG. BASH: a tool for managing BeadArray spatial artefacts. Bioinformatics (Oxford, England). 2008; 24(24):2921–2922.

[6] Silver JD, Ritchie ME, Smyth GK. Microarray background correction: maximum likelihood estimation for the normal-exponential convolution. Biostatistics. 2009 Apr;10(2):352-63.

[7] Smyth GK. Limma: Linear Models for Microarray Data. In R. Gentleman, V. Carey, S. Dudoit, and W. Huber R. Irizarry, editors, Bioinformatics and Computational Biology Solutions using R and Bioconductor. 2005; pp. 397–420. Springer, New York.

[8] Storey JD. The Positive False Discovery Rate: a Bayesian Interpretation and the q-value. The Annals of Statistics. 2003; 31(6):2013–2035.

[9] Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99(10):6567–6572.

[10] Dampney RA, Horiuchi J. Functional organisation of central cardiovascular pathways: studies using c-fos gene expression. Prog Neurobiol. 2003 Dec;71(5):359-84.

[11] Nitto T, Inoue T, Node K. Alternative spliced variants in the pantetheinase family of genes expressed in human neutrophils. Gene. 2008; 426(1-2):57–64.

[12] Martin F, Malergue F, Pitari G, Philippe JM, Philips S, Chabret C, Granjeaud S, Mattei MG, Mungall AJ, Naquet P, Galland F. Vanin genes are clustered (human 6q22-24 and mouse 10A2B1) and encode isoforms of pantetheinase ectoenzymes. Immunogenetics. 2001 May-Jun;53(4):296-306.

[13] Miyazaki K, Nagase T, Mesaki M, Narukawa J, Ohara O, Ishida N. Phosphorylation of clock protein PER1 regulates its circadian degradation in normal human fibroblasts. Biochem J. 2004 May 15;380(Pt 1):95-103.

[14] Bowden KL, Bilbey NJ, Bilawchuk LM, Boadu E, Sidhu R, Ory DS, Du H, Chan T, Francis GA. Lysosomal acid lipase deficiency impairs regulation of ABCA1 gene and formation of high density lipoproteins in cholesteryl ester storage disease. J Biol Chem. 2011 Sep 2;286(35):30624-35. Epub 2011 Jul 10.

[15] Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet. 2011 Mar 6;43(4):339-44.

[16] Wild PS, Zeller T, Schillert A, Szymczak S, Sinning CR, Deiseroth A, Schnabel RB, Lubos E, Keller T, Eleftheriadis MS, Bickel C, Rupprecht HJ, Wilde S, Rossmann H, Diemert P, Cupples LA, Perret C, Erdmann J, Stark K, Kleber ME, Epstein SE, Voight BF, Kuulaasma K, Li M, Schäfer AS, Klopp N, Braund PS, Sager HB, Demissie S, Proust C, König IR, Wichmann HE, Reinhard W, Hoffmann MM, Virtamo J, Burnett MS, Siscovick D, Wiklund PG, Qu L, El Mokthari NE, Thompson JR, Peters A, Smith AV, Yon E, Baumert J, Hengstenberg C, März W, Amouyel P, Devaney J, Schwartz SM, Saarela O, Mehta NN, Rubin D, Silander K, Hall AS, Ferrieres J, Harris TB, Melander O, Kee F, Hakonarson H, Schrezenmeir J, Gudnason V, Elosua R, Arveiler D, Evans A, Rader DJ, Illig T, Schreiber S, Bis JC, Altshuler D, Kavousi M, Witteman JC, Uitterlinden AG, Hofman A, Folsom AR, Barbalic M, Boerwinkle E, Kathiresan S, Reilly MP, O'Donnell CJ, Samani NJ, Schunkert H, Cambien F, Lackner KJ, Tiret L, Salomaa V, Munzel T, Ziegler A, Blankenberg S. A Genome-wide Association Study Identifies LIPA as a Susceptibility Gene for Coronary Artery Disease. Circ Cardiovasc Genet. 2011 May 23.

[17] Ligon KL, Fancy SP, Franklin RJ, Rowitch DH. Olig gene function in CNS development and disease. Glia. 2006 Jul;54(1):1-10.

[18] Wojakowski W, Ratajczak MZ, Tendera M. Mobilization of very small embryonic-like stem cells in acute coronary syndromes and stroke. Herz. 2010 Oct;35(7):467-72.

[19] Inouye M, Silander K, Hamalainen E, Salomaa V, Harald K, Jousilahti P, Männistö S, Eriksson JG, Saarela J, Ripatti S, Perola M, van Ommen G-JB, Taskinen M-R, Palotie A, Dermitzakis ET, Peltonen L. An Immune Response Network Associated with Blood Lipid Levels. PLoS Genet. 2010; 6(9).

[20] Takeda N, Maemura K, Imai Y, Harada T, Kawanami D, Nojiri T, Manabe I, Nagai R. Endothelial PAS domain protein 1 gene promotes angiogenesis through the transactivation of both vascular endothelial growth factor and its receptor, Flt-1. Circ Res. 2004; 95(2): 146-53.

[21] Castillo L, Rohatgi A, Ayers CR, Owens AW, Das SR, Khera A, McGuire DK, de Lemos JA. Associations of four circulating chemokines with multiple atherosclerosis phenotypes in a large population-based sample: results from the dallas heart study. J Interferon Cytokine Res. 2010 May;30(5):339-47.

[22] Chen Y, Xu H, Liu J, Zhang C, Leutz A, Mo X. The c-Myb functions as a downstream target of PDGF-mediated survival signal in vascular smooth muscle cells. Biochem Biophys Res Commun. 2007 Aug 24;360(2):433-6.

[23] Liang BT, Jacobson KA. A physiological role of the adenosine A3 receptor: sustained cardioprotection. Proc Natl Acad Sci U S A. 1998 Jun 9;95(12):6995-9.

[24] Liangos O, Domhan S, Schwager C, Zeier M, Huber PE, Addabbo F, Goligorsky MS, Hlatky L, Jaber BL, Abdollahi A. Whole blood transcriptomics in cardiac surgery identifies a gene regulatory network connecting ischemia reperfusion with systemic inflammation. PloS one. 2010; 5(10).

[25] Mallat Z, Corbaz A, Scoazec A, Besnard S, Les?che G, Chvatchko Y, Tedgui A. Expression of interleukin-18 in human atherosclerotic plaques and relation to plaque instability. Circulation. 2001 Oct 2;104(14):1598-603.

[26] Rosenberg S, Elashoff MR Beineke P Daniels SE, Wingrove JA,Tingley WG, Sager PT, Sehnert AJ, Yau M,Kraus WE, Newby LK, Schwarz RS,Voros S, Ellis SG, Tahirkheli N, Waksman R, McPherson J, Lansky A,Winn ME, Schork NJ,Topol EJ,for the PREDICT (Personalized Risk Evaluation and Diagnosis in the Coronary Tree) Investigators* Multicenter Validation of the Diagnostic Accuracy of a Blood-Based Gene Expression Test for Assessing Obstructive Coronary Artery Disease in Nondiabetic Patiens. Ann Intern Med. 2010;153:425-434.

[27] Tiret L, Godefroy T, Lubos E, Nicaud V, Tregouet DA, Barbaux S, Schnabel R, Bickel C, Espinola-Klein C, Poirier O, Perret C, Münzel T, Rupprecht HJ, Lackner K, Cambien F, Blankenberg S; AtheroGene Investigators. Genetic analysis of the interleukin-18 system highlights the role of the interleukin-18 gene in cardiovascular disease. Circulation. 2005 Aug 2;112(5):643-50.

[28] Uchida T, Nakao A, Nakano N, Kuramasu A, Saito H, Okumura K, Ra C, Ogawa H. Identification of Nash1, a novel protein containing a nuclear localization signal, a sterile alpha motif, and an SH3 domain preferentially expressed in mast cells. Biochem Biophys Res Commun. 2001 Oct 19;288(1):137-41.

[29] Lautier C, El Mkadem SA, Renard E, Brun JF, Gris JC, Bringer J, Grigorescu F. Complex haplotypes of IRS2 gene are associated with severe obesity and reveal heterogeneity in the effect of Gly1057Asp mutation. Hum Genet. 2003 Jul;113(1):34-43.

[30] Okazawa K, Yoshimasa Y, Miyamoto Y, Takahashi-Yasuno A, Miyawaki T, Masuzaki H, Hayashi T, Hosoda K, Inoue G, Nakao K. The haplotypes of the IRS-2 gene affect insulin sensitivity in Japanese patients with type 2 diabetes. Diabetes Res Clin Pract. 2005 Apr;68(1):39-48.

[31] Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki ML, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010 Oct 23;376(9750):1393-400.

[32] Lluís-Ganella C, Lucas G, Subirana I, Sentí M, Jimenez-Conde J, Marrugat J, Tomás M, Elosua R. Additive effect of multiple genetic variants on the risk of coronary artery disease. Rev Esp Cardiol. 2010 Aug;63(8):925-33.

[33] Browman DT, Resek ME, Zajchowski LD, Robbins SM. Erlin-1 and erlin-2 are novel members of the prohibitin family of proteins that define lipid-raft-like domains of the ER. J Cell Sci. 2006 Aug 1;119(Pt 15):3149-60.

[34] Gillespie CF, Phifer J, Bradley B, Ressler KJ. Risk and resilience: genetic and environmental influences on development of the stress response. Depress Anxiety. 2009; 26(11): 984-92. Review.

[35] Zhang SX, Garcia-Gras E, Wycuff DR, Marriot SJ, Kadeer N, Yu W, Olson EN, Garry DJ, Parmacek MS, Schwartz RJ. Identification of direct serum-response factor gene targets during Me2SO-induced P19 cardiac cell differentiation. J Biol Chem. 2005 May 13;280(19):19115-26.

[36] Liu C, Shi Y, Du Y, Ning X, Liu N, Huang D, Liang J, Xue Y, Fan D. Dual-specificity phosphatase DUSP1 protects overactivation of hypoxia-inducible factor 1 through inactivating ERK MAPK. Exp Cell Res. 2005 Oct 1;309(2):410-8.

[37] Dumont J, Zureik M, Cottel D, Montaye M, Ducimeti?re P, Amouyel P, Brousseau T. Association of arginase 1 gene polymorphisms with the risk of myocardial infarction and common carotid intima media thickness. J Med Genet. 2007 Aug;44(8):526-31.

[38] Harpster MH, Bandyopadhyay S, Thomas DP, Ivanov PS, Keele JA, Pineguina N, Gao B, Amarendran V, Gomelsky M, McCormick RJ, Stayton MM.Earliest changes in the left ventricular transcriptome postmyocardial infarction. Mamm Genome. 2006 Jul;17(7):701- 15.

[39] Acosta J, Hettinga J, Flückiger R, Krumrei N, Goldfine A, Angarita L, Halperin J. Molecular basis for a link between complement and the vascular complications of diabetes. Proc Natl Acad Sci USA. 2000 May 9;97(10):5450-5.

[40] Wu G, Hu W, Shahsafaei A, Song W, Dobarro M, Sukhova GK, Bronson RR, Shi GP, Rother RP, Halperin JA, Qin X. Complement Regulator CD59 Protects Against Atherosclerosis by Restricting the Formation of Complement Membrane Attack Complex Circulation Research 2009, 104:550-558.

[41] Cheng C, Tempel D, Den Dekker WK, Haasdijk R, Chrifi I, Bos FL, Wagtmans K, van de Kamp EH, Blonden L, Biessen EA, Moll F, Pasterkamp G, Serruys PW, Schulte-Merker S, Duckers HJ. Ets2 determines the inflammatory state of endothelial cells in advanced atherosclerotic lesions. Circ Res. 2011 Aug 5;109(4):382-95.

[42] Eijgelsheim M, Newton-Cheh C, Sotoodehnia N, de Bakker PI, Müller M, Morrison AC, Smith AV, Isaacs A, Sanna S, Dörr M, Navarro P, Fuchsberger C, Nolte IM, de Geus EJ, Estrada K, Hwang SJ, Bis JC, Rückert IM, Alonso A, Launer LJ, Hottenga JJ, Rivadeneira F, Noseworthy PA, Rice KM, Perz S, Arking DE, Spector TD, Kors JA, Aulchenko YS, Tarasov KV, Homuth G, Wild SH, Marroni F, Gieger C, Licht CM, Prineas RJ, Hofman A, Rotter JI, Hicks AA, Ernst F, Najjar SS, Wright AF, Peters A, Fox ER, Oostra BA, Kroemer HK, Couper D, Völzke H, Campbell H, Meitinger T, Uda M, Witteman JC, Psaty BM, Wichmann HE, Harris TB, Kääb S, Siscovick DS, Jamshidi Y, Uitterlinden AG, Folsom AR, Larson MG, Wilson JF, Penninx BW, Snieder H, Pramstaller PP, van Duijn CM, Lakatta EG, Felix SB, Gudnason V, Pfeufer A, Heckbert SR, Stricker BH, Boerwinkle E, O'Donnell CJ. Genome-wide association analysis identifies multiple loci related to resting heart rate. Hum Mol Genet. 2010 Oct 1;19(19):3885-94.

[43] Puigdecanet E, Espinet B, Lozano JJ, Sumoy L, Bellosillo B, Arenillas L, Alvarez-Larrán A, Solé F, Serrano S, Besses C, Florensa L. Gene expression profiling distinguishes JAK2V617F-negative from JAK2V617F-positive patients in essential thrombocythemia. Leukemia. 2008 Jul;22(7):1368-76.

[44] Valenta, Z. Genetic determinants of AMI: Supplemental Material. [online]. 2012 [cited 2012 Dec 30]. Available from: http://www.ejbi.org/img/ejbi/2012/1/att/Valenta.pdf