Potential biomarkers of acute myocardial infarction based on weighted gene co-expression network analysis

Zhihua Liu[1,3,4,5*], Chenguang Ma[2,3,4,5], Junhua Gu[4,5,6] and Ming Yu[4,5,6]

*Correspondence:
zh.liu@siat.ac.cn
[1] Shenzhen Institutes
of Advanced
Technology, Chinese
Academy of Sciences,
Shenzhen 518055, China
Full list of author information
is available at the end of the
article

Abstract

Background: Acute myocardial infarction (AMI) is the common cause of mortality in developed countries. The feasibility of whole-genome gene expression analysis to identify outcome-related genes and dysregulated pathways remains unknown. Molecular marker such as *BNP*, *CRP* and other serum inflammatory markers have got the notice at this point. However, these biomarkers exhibit elevated levels in patients with thyroid disease, renal failure and congestive heart failure. In this study, three groups of microarray data sets (GES66360, GSE48060, GSE29532) were collected from GEO, a total of 99, 52 and 55 samples, respectively. Weighted gene co-expression network analysis (WGCNA) was performed to obtain a classifier which composed of related genes that best characterize the AMI.

Results: Here, this study obtained three groups of microarray data sets (GES66360, GSE48060, GSE29532) on AMI blood samples, a total of 99, 52 and 24 samples, respectively. In all, 4672 genes, 3185 genes, 3660 genes were identified in GSE66360, GSE48060, GSE60993 modules, respectively. We preformed WGCNA, GO and KEGG pathway enrichment analysis on these three data sets, finding function enrichment of the differential expression gene on inflammation and immune response. Transcriptome analysis were performed in AMI patients at four time points compared to CAD patients with no history of MI, to determine gene expression profiles and their possible changes during the recovery from myocardial infarction.

Conclusions: The results suggested that three overlapping genes (*FGFBP2*, *GFOD1* and *MLC1*) between two modules could be a potential use of gene biomarkers for the diagnose of AMI.

Keywords: Acute myocardial infarction, Weighted gene co-expression network analysis, Gene ontology, Functional enrichment analysis

Introduction

Despite significant advances in revascularization strategies, pharmacotherapy, cardiac rehabilitation algorithms and organ transplantation, cardiovascular diseases, including acute myocardial infarction (AMI), remains the leading cause of death in developed countries [1]. The classic risk factors such as smoking, high serum cholesterol, hypertension and diabetes mellitus that can partly predict in disease prevention and outcomes, but is not sufficient to fully provide acute diagnosis [2]. An important challenge

to implementing such strategies is the limited predictive value of current risk forecast models. This "detection gap" is illustrated by the observation that patients with coronary heart disease often lack conventional risk factors, up to 20% of patients have no traditional risk factors, and 40% have only one [3, 4]. Currently more attention is given to nontraditional as well as genetics risk factors, thus, improved strategies for controlling the development of AMI, reducing the mortality rate and improving prognosis are a public health priority.

Molecular marker such as *BNP*, *CRP* and other serum inflammatory markers have got the notice at this point, however, these biomarkers exhibit elevated levels also in patients with thyroid disease, renal failure and congestive heart failure [5], mandating the search for novel more sensitive risk markers to improve the selection of individuals for preventative strategies. Such biomarkers must meet the ability to identify individuals at risk, stability of results when repeated and therapeutic impact with early intervention, and motivate hypotheses for cell-based therapeutic targets to control the pathological process of AMI.

Invasive coronary angiography is the "gold standard" for detecting AMI. However, this is costly, and can pose risk to the patient. Genome-wide expression analysis using microarray is an extensively used strategy for the detection of new biomarkers for diagnosis, prediction of disease severity and identification of novel drug targets [6, 7]. In this study, our main goal were: (1) to establish alterations in gene expression patterns in a peripheral blood cell model from patients with AMI and through a follow up; (2) to identify distinct biomarkers that correlate with AMI and development.

For co-expression data analysis, three groups of microarray data sets (GES66360, GSE48060, GSE29532) were downloaded from GEO, a total of 99, 52 and 55 samples, respectively. We proposed a simple analysis method used for finding clusters (modules) of highly correlated genes, called "weighted correlation network analysis" (WGCNA) to obtain a classifier which composed of related genes that best characterize the AMI. In addition, we performed gene ontology (GO) and Kyoto encyclopaedia of genes and genomes (KEGG) pathway analyses of the significant clusters.

Method

Data

The microarray expression data sets (GES66360, GSE48060, GSE60993) were downloaded from the GEO database. This study samples consisted of first-time AMI patients and healthy controls with a normal echocardiogram. The data set GSE60993 was collected from peripheral blood containing 17 patients and 7 control samples, using the Illumina HumanWG-6 v3.0 expression beadchip. The other two data sets GSE48060 and GES66360 were both used the Affymetrix Human Genome U133 Plus 2.0 Array platform, including 49 patients versus 50 controls collected from circulating endothelial cells, and 31 patients versus 21 controls collected from peripheral blood, respectively.

The aim of our second microarray cohort analysis was to find expression of genes linked specially with the follow-up of AMI,to exclude genes linked with coronary artery disease (CAD). Here, we chose the control group patients with stable CAD, not the healthy controls.

Differential expression

Before the differential expression analysis, we performed the step of preprocessing. If a gene had more than one probe site, we averaged the values of the probe sites.

After the preprocessing, we performed variance analysis between the AMI patient group and the healthy control group for the three data sets. And the statistical significance threshold level for differential expression genes was $P < 0.05$. So we identified 4672 genes, 3185 genes, 1167 genes in GES66360, GSE48060, GSE29532 data sets, respectively.

Weighted gene co-expression network analysis (WGCNA)

WGCNA was used the scale free topology criterion to construct the gene co-expression network. The differential expression genes of the three data sets were independently analyzed using the WGCNA method. First, for each set of genes, a weighted correlation network was constructed by computing a pair-wise correlation matrix. We arrange the correlations into a matrix: Let p_{ij} be the correlation between variables i and j. Place p_{ij} into positions (i, j) and (j, i) of the correlation matrix. The absolute value of the correlation is used as an unsigned co-expression similarity measure (Eq. 1).

$$p_{ij}^{unsigned} = \left| cor\left(x_i,\ x_j\right) \right|. \tag{1}$$

Next, an adjacency matrix was constructed by raising the correlation matrix to the power of 8. Based on the adjacency matrix, the topological overlap measure of each pair of genes was calculated. Then, we performed average linkage hierarchical clustering on the topological overlap, and used the Dynamic Hybrid Tree Cut algorithm to cut the clustering tree. To obtain large and distinct modules, we set the minimum modules size to 30 genes and the minimum height for merging modules at 0.2.

Functional enrichment analysis

For each module genes, GO and Pathway Functional enrichment analysis was assessed using hypergeometric test. The calculating formula is:
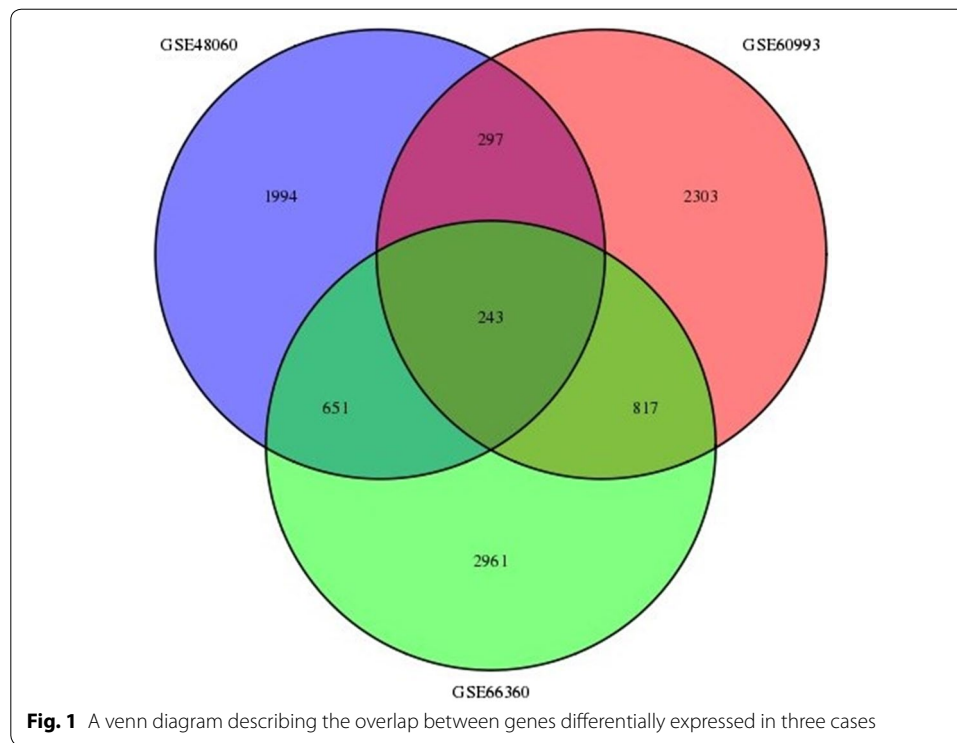
$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}. \tag{2}$$

In Eq. (2), N is the number of all genes with annotation; n is the number of module genes in N; M is the number of all genes that are annotated to the certain terms; m is the number of module genes in M. The calculated P value goes through Bonferroni Correction, taking corrected P value 0.05 as a threshold.

Results

Microarray differentially expressed gene between AMI and healthy controls

We first selected the global different genes in the organization of gene transcriptome between AMI and control groups. We identify 4672 genes, 3185 genes, 3660 genes

Fig. 1 A venn diagram describing the overlap between genes differentially expressed in three cases

Table 1 The pathways involved in three cases

Modules	Pathways	Q-value (Q < 0.05)
GSE48060	Natural killer cell mediated cytotoxicity	1.066849e−07
	Graft-versus-host disease	1.956196e−07
	Allograft rejection	1.237023e−05
	Cytokine-cytokine receptor interaction	1.231519e−04
	Type I diabetes mellitus	3.664642e−04
GSE60993	Ribosome	1.315470e−06
	Intestinal immune network for IgA production	8.294981e−05
	Allograft rejection	1.018873e−04
	Hematopoietic cell lineage	3.178231e−04
	NF-kappa B signaling pathway	4.203928e−04
GSE66360	Graft-versus-host disease	7.291560e−18
	Natural killer cell mediated cytotoxicity	1.330539e−15
	Antigen processing and presentation	1.330539e−15
	Cytokine-cytokine receptor interaction	3.994169e−01
	Gap junction	3.994169e−01

in GSE66360, GSE48060, GSE60993 groups, respectively (Fig. 1). The significant expression changes could be showed between AMI and control groups. We observed a highly significant overlap between differentially expressed genes in three of cases, supporting the robustness of the data and indicating the AMI-specific expression changes are consistent across these three cases.

Liu *et al. BioMed Eng OnLine*   (2019) 18:9

Page 5 of 12

(See figure on next page.)
Fig. 2 The heat map of modules correspond to GSE48060 (**a**), GSE60993 (**b**), GSE66360 (**c**) differentially expressed between AMI and control samples. Scaled expression values are colour-coded according to the legend on the left. The dendrogram depicts hierarchical clustering of samples and gene expression based on the differentially expressed genes. The top bar indicates the disease status: red, AMI; black, control

Next, we applied WGCNA to integrate the expression differences into a higher order, systems level context. The pathways involved in each module were summarized in Table 1.

We chose the appropriate module which show significant differences (Q < 0.05) and related to the progression of AMI. For these three sets of data, we selected the midnight-blue module of GSE66360 group, the pink module of the GSE48060 group, the brown module of the GSE60993 group. Then supervised hierarchical clustering based on the top differentially expressed genes showed distinct clustering of the samples.
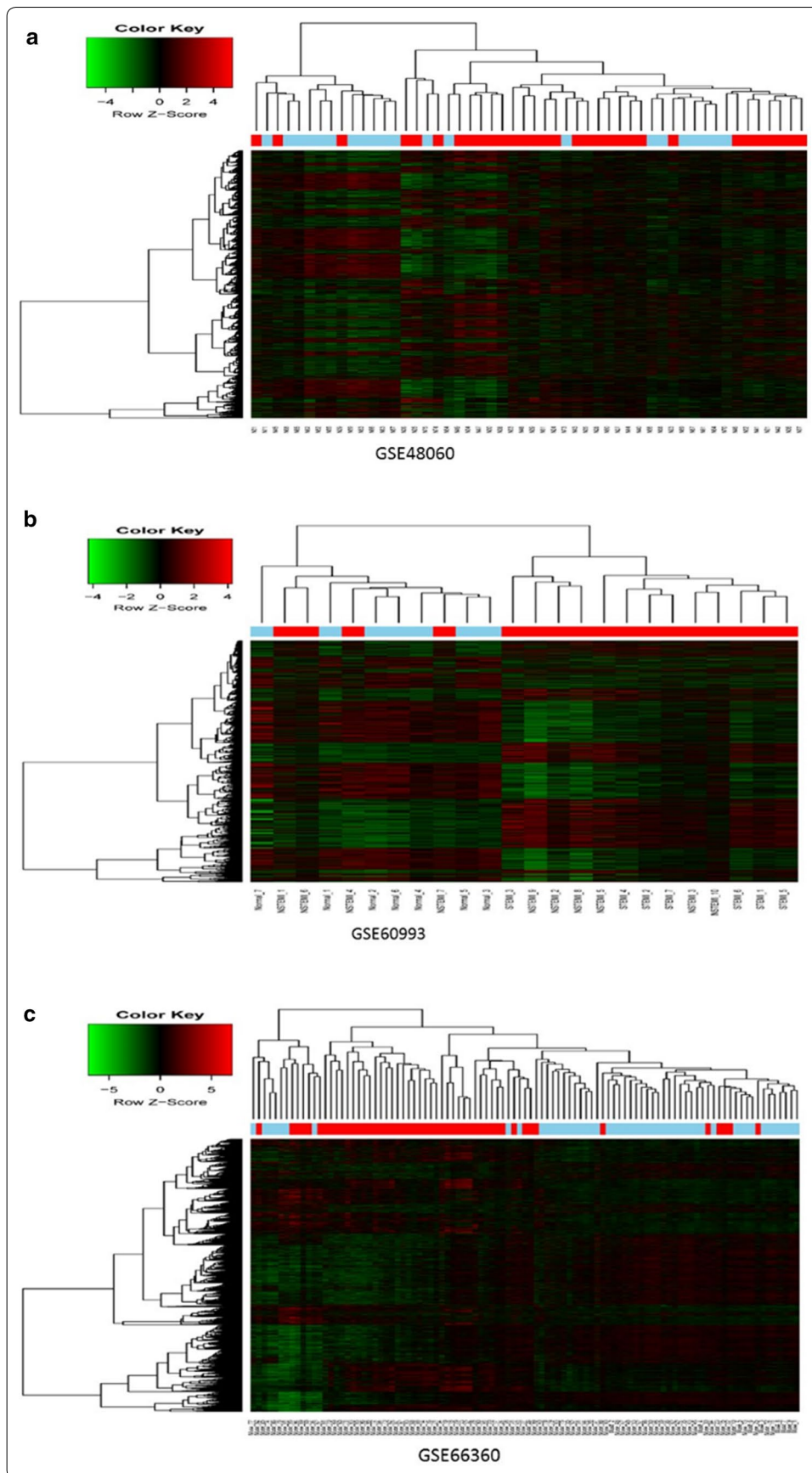
Clustering analysis did not identify significant global differences between groups on visual inspection. However, hierarchical cluster analysis showed clustering of gene expression and samples in the GSE66360, GSE48060, GSE60993 group, all three were divided into two distinct parts. Genes involved in the differential module were then subjected to functional annotation and pathway enrichment analysis, the specific results as shown in the Fig. 2a–c. We also observed only three overlapping genes between differentially expressed genes in these three selected modules: *FGFBP2*, *GFOD1*, *MLC1*, which were functionally enriched for revascularization, glycometabolism, Megalencephalic leukoencephalopathy with subcortical cysts pathways.

Gene expression profiling at different time points following AMI

To determine gene expression profiles and their possible changes during the recovery from myocardial infarction, we performed a transcriptome analysis in AMI patients at four time points compared to CAD patients with no history of MI.

A comparison between the samples from the first two time points (admission, discharge after AMI) and samples from the same patients collected 6 months after AMI or control group (patients with a stable coronary artery disease and without a history of myocardial infarction) was performed to identify genes between the comparisons. We also applied WGCNA to identify modules of co-expressed genes with high topological overlap. Finally, module eigengenes are calculated, which provides quantitative assessments of modules for further analysis. We identified a 172 transcripts overlap between the GSE59867 and GSE62646 data set on admission versus the control group (Fig. 3).

In a canonical analysis three most significant pathways were: Natural killer cell mediated cytotoxicity, Graft-versus-host disease, Antigen processing and presentation (Table 2). The expression of 427 genes within 6 months after AMI was changed relative to admission, and pathway analysis of these different expressed transcripts were mainly associated with Oxidative phosphorylation, Huntington's disease and Parkinson's disease. On discharge 157 transcripts were differentially expressed compared to control group, and 179 transcripts compared to 6 months after AMI. Notably, the pathway enrichment analysis results showed that admission versus control and discharge versus
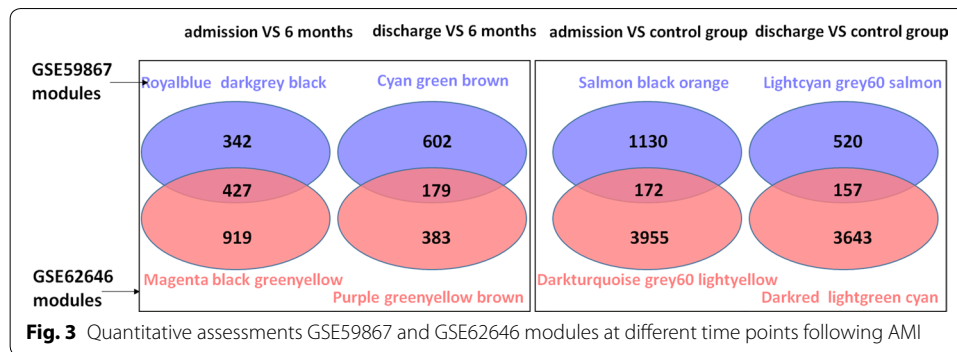
Liu *et al. BioMed Eng OnLine*   (2019) 18:9

Page 7 of 12

Fig. 3 Quantitative assessments GSE59867 and GSE62646 modules at different time points following AMI

Table 2 The pathways involved in differentially expressed genes compared with different time points following AMI

	Pathways	Q-value (Q < 0.05)
admission versus control group	Natural killer cell mediated cytotoxicity	3.142005e−09
	Graft-versus-host disease	5.779472e−07
	Antigen processing and presentation	2.997803e−03
	Cytosolic DNA-sensing pathway	2.302012e−02
discharge versus control group	Natural killer cell mediated cytotoxicity	8.154634e−10
	Graft-versus-host disease	1.164728e−08
	Antigen processing and presentation	1.657503e−04
	Allograft rejection	2.438676e−02
admission versus 6 months	Oxidative phosphorylation	3.841450e−15
	Huntington's disease	1.570003e−12
	Parkinson's disease	2.462576e−12
	Alzheimer's disease	1.076534e−11
discharge versus 6 months	Parkinson's disease	4.627948e−16
	Oxidative phosphorylation	4.627948e−16
	Alzheimer's disease	9.071314e−15
	Huntington's disease	8.428429e−14

control group share the similar pathways, the admission versus 6 month and discharge versus 6 month group share the similar pathways.

Additionally, we undertook a patient-by-patient analysis of the most differently expressed genes between above four groups, and identified 19 common genes in all patients. The direction of the most significant change was the same although the relative levels of expression differed markedly between patients.

Discussions

The most important limitation for accurate investigation of the pathophysiology of AMI is the necessity for heart tissue sample [8, 9]. Transcriptome technologies have provided new opportunities to discover disease-specific mechanisms which were distinguishable and may offer diagnostic and prognostic value [10–18]. In this paper, we compared the two transcriptome profiles: AMI patients versus healthy controls, admission, discharge, and 6 months after AMI patients versus with the controls (not the healthy group), with

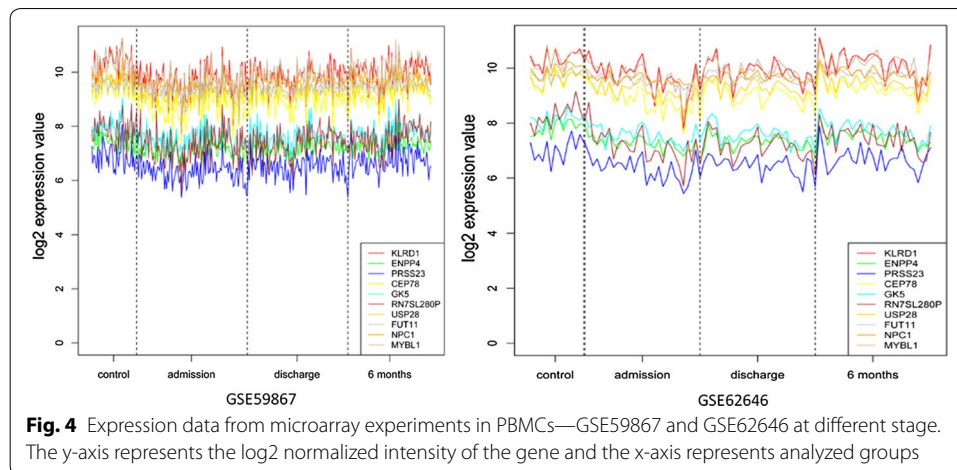Liu *et al. BioMed Eng OnLine*   (2019) 18:9

Page 8 of 12

the aims to identify potential cardiac ischemia-related biomarkers and revealed the anticipated modulation networks and pathways that correlate with AMI development.

AMI patients versus healthy controls

We retained the 243 overlap between differentially expressed genes in AMI and control group in three cases, which were functionally enriched for inflammation, immune response pathways in AMI patients compared to matched, healthy controls. These findings are consistent with previous studies which also demonstrated the increased gene levels for immune response, inflammation and apoptosis pathways, raising the possibility of relationship between myocardial pathological procedure and inflammatory transcriptional changes in circulating cells. Indeed, targeted gene studies by measuring directly the inflammatory mRNA profile of leucocytes in multigene system, supporting the hypothesis that an inflammatory response involving leucocytes contributed to the pathogenesis of myocardial infarction [19]. For another, as in the modulation of leukocyte transcriptional machinery, the analyses of changes in blood leukocyte gene expression patterns reveals that the human blood leukocyte response to transient dysregulation and modulation of translational machinery, providing further evidence that myocardial ischemia results in transcriptional gene expression changes in the peripheral blood [20, 21].

The complexity of the dataset was reduced by lowing non-significant probe sets using WGCNA method, the most significant module was generated in each case. We found only three overlapping genes between the pink module of the GSE48060 group, the midnightblue module of the GSE66360 group and the brown module of the GSE60993 group. These three genes were FGFBP2, GFOD1 and MLC1. This finding implied the AMI pathogeny was associated with an increased expression of genes involved in the intravascular lesions, immunological responses and brain-derived factor regulatory system, suggesting a potential use of FGFBP2, GFOD1 and MLC1 as gene expression biomarkers for very early stages of AMI.

We then selected the co-expression module that showed great significance to analyze biological functions, these analyses associated inflammation and immune response pathway including natural killer cell mediated cytotoxicity, graft-versus-host disease, allograft rejection, antigen processing and presentation and Cytokine-cytokine receptor interaction with pathological changes following ischemic cardiac injury. These findings provide evidence that profiling of circulating cells in patients is capable of identifying distinct, biologically relevant gene expression changes and modulated networks and pathways. It's worth mentioning that functional analysis of these differential gene sets revealed the enrichment of Type I diabetes mellitus and Autoimmune thyroid disease pathways. The subclinical hypothyroidism (associated with reduced systolic function, an atherogenic lipid profile, diastolic hypertension, and inflammatory condition) and hyperthyroidism (related to a mild decreased of coronary reserve, and an increased risk of supraventricular arrhythmias, hypercoagulable state) have recently been documented as clinical entities with negative effects on the cardiovascular system [22, 23]. Thus, mild forms of thyroid hormones disorders, often leads to the similar changes in cardiac function and gene expression, even small variations of the thyroid hormone within the

Fig. 4 Expression data from microarray experiments in PBMCs—GSE59867 and GSE62646 at different stage. The y-axis represents the log2 normalized intensity of the gene and the x-axis represents analyzed groups
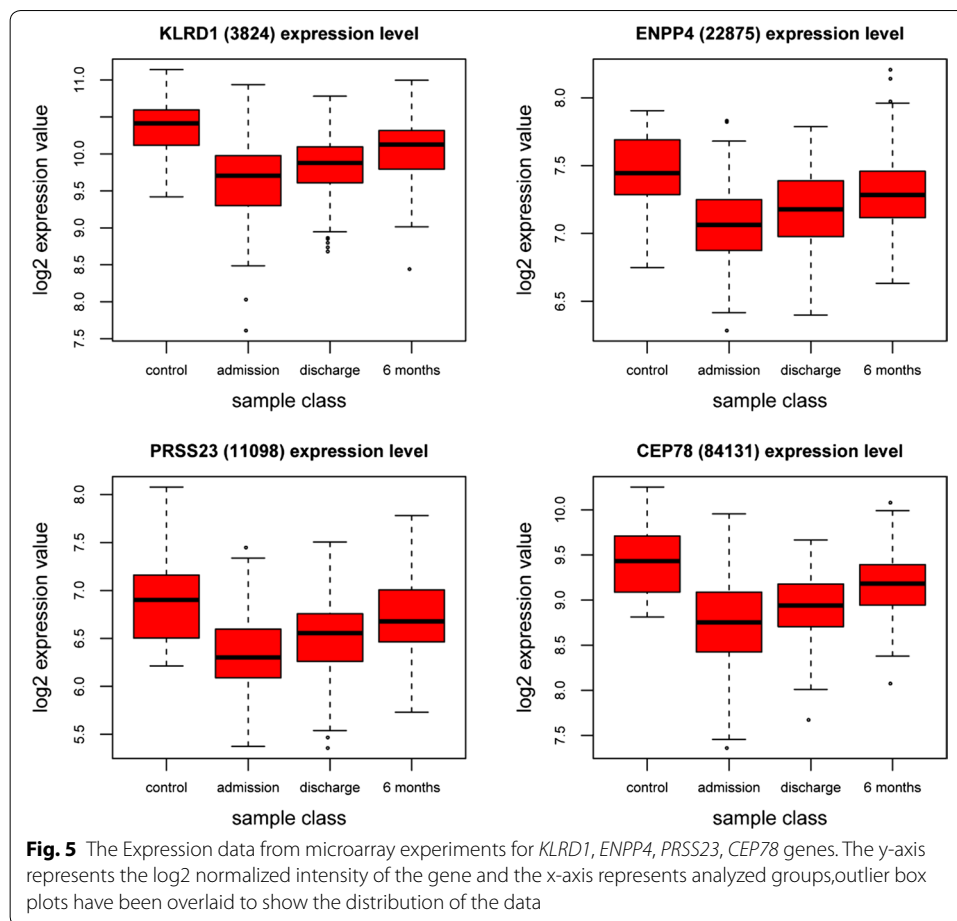
physiological range may cause the adverse cardiovascular development. This supports the utility of large-scale gene expression analysis of the blood transcriptome at initial diagnosis of AMI for identification of relevant mechanisms of disease.

Admission, discharge, and 6 months after AMI patients versus with the controls

Numerous studies found significant roles of PBMCs in the systemic and regional inflammatory responses associated with remodeling in AMI [24, 25]. Thus, activation of the PBMCs, which reflects the magnitude of inflammation, could be linked to the progression of AMI patients.

The prognosis of patients after acute coronary syndrome (ACS) largely depends on the extent of myocardial damage during the acute phase. In this study, two gene expression profiling in PBMCs—GSE59867 and GSE62646 were downloaded to identify biologically relevant transcripts significantly altered through the AMI follow-up. Next, we performed four comparison: admission versus control, discharge versus control, admission versus 6 month and discharge versus 6 month group, respectively (Fig. 4). Notably, most differentially expressed genes were common in admission versus control and discharge versus control group, with maximum differentially expressed genes in admission versus 6 month and discharge versus 6 month group respectively, which is the same in the pathway analysis (Fig. 5).

These indicating that these gene expression profile modification may be responsible for the functional change during the recovery from AMI. Consistent with pathological changes following ischemic cardiac injury, the top five pathways enriched in the comparison of admission or discharge versus control group were associated with inflammation and immune response, which is also consistent with the previous results in the first section. Interestingly, possibly duo to the pharmacological intervention and myocardial function recovery, the comparison of admission or discharge versus 6 month were functional enriched in metabolic pathways and neurological disease. One possibility is that neurosecretion was involved in a protective systemic response to the development of the cardiac insufficiency in the face of partial or total occlusion of the coronary artery. Previous studies [26] have demonstrated that tyrosine kinase

Fig. 5 The Expression data from microarray experiments for *KLRD1*, *ENPP4*, *PRSS23*, *CEP78* genes. The y-axis represents the log2 normalized intensity of the gene and the x-axis represents analyzed groups,outlier box plots have been overlaid to show the distribution of the data

receptor B (TrkB)—a high-affinity receptor for brain-derived neurotrophic factor protects endothelial integrity during atherogenesis and plays a previously unknown protective role in the development of CAD. Also, the long term activation of autonomic nervous system may increase the risk of atherosclerosis, myocardial infarction or sudden death [27, 28].

Conclusion

In conclusion, the transcriptional profiling presented here demonstrating relationship between inflammation and immune response, Type I diabetes and thyroid disorder and AMI pathological process. On the other hand, we found a set of genes involved in inflammation and immune response on admission and discharge relative to the stable coronary artery disease, and the follow-up of AMI to 6 months associated with the metabolic pathways or neurosecretion. This pilot study warrants further investigation with larger cohorts in the setting of coronary heart disease, it should be very interesting to determine possible interferences in evaluation of diagnostic sensitivity and specificity of the corresponding marker genes, so was the prognostic applicability.

Author details
^{1} Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. ^{2} Tsinghua University, Beijing 100084, China. ^{3} Beijing Yuqiu Medical Research Institute, Beijing 100022, China. ^{4} Shenzhen Yuqiu Biological Big Data Research Institute, Shenzhen 518033, China. ^{5} Nanjing Yuqiu Biotechnology Co., Ltd., Nanjing 210009, China. ^{6} Hebei University of Technology, Tianjin 300130, China.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Mozaffarian D, et al. Heart disease and stroke statistics-2016 update: a report from the American heart association. Circulation. 2016;133(4):e38–60.
2. Law MR, Watt HC, Wald NJ. The underlying risk of death after myocardial infarction in the absence of treatment. Arch Intern Med. 2002;162(21):2405–10.
3. Ajani UA, Ford ES. Has the risk for coronary heart disease changed among US adults? J Am Coll Cardiol. 2006;48(6):1177–82.
4. Hozawa A, et al. Absolute and attributable risks of cardiovascular disease incidence in relation to optimal and borderline risk factors: comparison of African American with white subjects–atherosclerosis risk in communities study. Arch Intern Med. 2007;167(6):573–9.
5. Ge Y, Wang TJ. Identifying novel biomarkers for cardiovascular disease risk prediction. J Intern Med. 2012;272(5):430–9.
6. Hall PA, et al. An introduction to genes, genomes and disease. J Pathol. 2010;220(2):109–13.
7. Heidecker B, Hare JM. The use of transcriptomic biomarkers for personalized medicine. Heart Fail Rev. 2007;12(1):1–11.
8. Devaux Y, et al. Transforming growth factor beta receptor 1 is a new candidate prognostic biomarker after acute myocardial infarction. BMC Med Genom. 2011;4:83.
9. Kim J, et al. Gene expression profiles associated with acute myocardial infarction and risk of cardiovascular death. Genome Med. 2014;6(5):40.
10. Liu ZH, Yang D, Xie P, et al. MiR-106b and MiR-15b modulate apoptosis and angiogenesis in myocardial infarction. Cell Physiol Biochem. 2012;29:851–62.
11. Ren GM, Liu ZH. NetCAD: a network analysis tool for coronary artery disease associated PPI network. Bioinformatics. 2013;29(2):279–80.
12. Yang D, Xie P, Liu ZH. Ischemia/reperfusion-induced MKP-3 impairs endothelial NO formation via inactivation of ERK1/2 pathway. PLoS ONE. 2012;7(7):e42076.
13. Cheng JK, Cao FL, Liu ZH. AGP: a multimethods web server for alignment-free genome phylogeny. Mol Biol Evol. 2013;30(5):1032–7.
14. Liu ZH, Chu GY. Chronobiology in mammalian health. Mol Biol Rep. 2013;40:2491–501.
15. Wei L, Wan S, Guo J, Wong KKL. Novel hierarchical selected ensemble classifier with bioinformatics application. Artif Intell Med. 2017;83:82–90.
16. Wang D, Fong S, Wong RK, Mohammed S, Fiaidhi J, Wong KKL. Robust high-dimensional bioinformatics data streams mining by ODR-ioVFDT. Sci Rep. 2017;7:43167.
17. Li J, Fong S, Wong RK, Millham R, Wong KKL. Elitist binary wolf search algorithm for feature selection in high-dimensional bioinformatics datasets. Sci Rep. 2017;7:4354.
18. Li J, Fong S, Sung Y, Cho K, Wong R, Wong KKL. Adaptive swarm cluster-based dynamic multi-objective synthetic minority oversampling technique algorithm for tackling binary imbalanced datasets in biomedical classification. BioData Min. 2016;9:37.
19. Suresh R, et al. Transcriptome from circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction. J Mol Cell Cardiol. 2014;74:13–21.
20. Calvano SE, et al. A network-based analysis of systemic inflammation in humans. Nature. 2005;437(7061):1032–7.
21. Wettinger SB, et al. High throughput mRNA profiling highlights associations between myocardial infarction and aberrant expression of inflammatory molecules in blood cells. Blood. 2005;105(5):2000–6.

22. Abdulaziz Qari F. Thyroid hormone profile in patients with acute coronary syndrome. Iran Red Crescent Med J. 2015;17(7):e26919.
23. Cokkinos DV, Chryssanthopoulos S. Thyroid hormones and cardiac remodeling. Heart Fail Rev. 2016;21(4):365–72.
24. Bauters A, et al. Relation of admission white blood cell count to left ventricular remodeling after anterior wall acute myocardial infarction. Am J Cardiol. 2007;100(2):182–4.
25. Aoki S, et al. Elevated peripheral blood mononuclear cell count is an independent predictor of left ventricular remodeling in patients with acute myocardial infarction. J Cardiol. 2011;57(2):202–7.
26. Jiang H, et al. Tyrosine kinase receptor B protects against coronary artery disease and promotes adult vasculature integrity by regulating Ets1-mediated VE-cadherin expression. Arterioscler Thromb Vasc Biol. 2015;35(3):580–8.
27. Ozturk S, et al. Anxiety and depression scores in patients with coronary artery disease and coronary artery ectasia. Int J Cardiol. 2015;186:299–301.
28. Hausberg M, Hillebrand U, Kisters K. Addressing sympathetic overactivity in major depressive disorder. J Hypertens. 2007;25(10):2004–5.