Potential Role of a Three-Gene Signature for Predicting Diagnosis in Patients With Myocardial Infarction

Yinhui Yao
the affiliated of Chengde medical college https://orcid.org/0000-0002-1244-0930
Jingyi Zhao (33236986@qq.com)
Chengde Medical College. Shangerdao hezi
Xiaohui Zhou
Chengde medical college
Ying Wang
the affiliated of chengde medical college

Research

Keywords: Gene signature, Myocardial infarction, Diagnosis, Meta-analysis, Biomarker

Posted Date: January 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-138899/v1

Version of Record: A version of this preprint was published at Bioengineered on January 1st, 2021. See the published version at https://doi.org/10.1080/21655979.2021.1938498.

Abstract

Objective

Increasing evidence underscored that the expression of genes was associated with the development and progression of myocardial infarction (MI). In this study, We evaluated the diagnostic value of the feature genes in MI based on data from the Gene Expression Omnibus (GEO) database.

Methods

We used the data from the GEO database (GSE66360) to identify a set of significant differentially expressed genes (DEGs) between MI and healthy control. Univariable logistic regression, the least absolute shrinkage and selection operator (LASSO), SignalP 3.0 server and multivariable logistic regression were used to find the potential role of genes for predicting diagnosis in patients with MI. Receiver operating characteristic (ROC) curve analyses, area under the curve (AUC) and C-index were used to estimate the diagnostic value of genes in patients with MI. The validation for the association was conduct in another six independent data sets (GSE141512, GSE24519, GSE34198, GSE48060, GSE60993, and GSE109048). Then, a meta-analysis was performed to evaluate the diagnostic value of genes in MI.

Results

A total of 44 DEGs were selected from GSE66360. Functional enrichment and KEGG analysis were performed to reveal the DEGs in some inflammation-related biological processes and pathways. A three-gene signature consisted of CCL20, IL1R2 and ITLN1, which could effectively distinguish patients in MI (AUC and C-index were the same value of 0.975). The three-gene signature was effectively validated in 7 independent cohorts, and diagnostic meta-analysis results of the three-gene signature showed that the pooled sensitivity, specificity and ROC curve AUC for MI were0.82 (95% CI: 0.68-0.90), 0.91 (95% CI: 0.81-0.96) and 0.94(95%CI, 0.91-0.96), respectively.

Conclusion

It was magnificently suggest that the three-gene signature might potentially serve as novel candidate biomarkers for distinguishing MI from healthy control. Besides, more well-designed cohort studies need to be implemented to warrant the diagnostic value of three-gene signature in clinical purpose.

1. Introduction

Myocardial infarction (MI), also known as a heart attack, is one of the leading causes of hospital admission and mortality worldwide (WANG and JING 2018). Early prevention, screening, monitoring, diagnosis and treatment may reduce the incidence and mortality of MI. However, the recent of research advances in effective treatment for MI was still lacking, so the best strategies for a more important method focused on early diagnosis aiming at managing the underlying etiologies and complications of MI. Although cardiac troponin T (cTnT) and creatine kinase MB (CK-MB) for MI are useful of diagnostic

tools, there a relatively low diagnostic accuracy limit their applications (DE WINTER et al. 1995; LIU et al. 2018; ZHAO et al. 2019). Previous studies also showed that a relatively low level of cTnT was challenging to detect in healthy human serum (CHRISTENSON et al. 2000; CHAN and NG 2010). The concentration of CK-MB in the blood decreased gradually after the onset of acute MI 36–72 h, which was almost equivalent normal levels(CHRISTENSON et al. 2000; RAKOWSKI et al. 2014). Furthermore, molecular markers are critical for the research and clinical treatment of cardiovascular diseases (PARK et al. 2015; CHEN et al. 2019; GOBBI et al. 2019). Therefore, identifying the promising novel molecular markers is critical demanded, which will contribute to enhance our understanding of MI initiation and progression and promote early detection of MI.

At present, The National Center for Biotechnology Information developed the Gene Expression Omnibus (GEO) database, which was a consolidation of available transcriptomic data for further expanding the scope of biomedical research. With the quick development of gene microarray technology, it provided an efficient alternative for screening genetic alterations at the genome level, which was beneficial for us to confirm the differentially expressed genes (DEGs) and functional pathways involved the progression of MI. However, it was challenging to identify reliable results that were conducted by independent microarray analysis. Many studies reported that the novel molecular markers were to identify for predicting diagnosis in patients with MI and underlying the mechanisms of MI by using microarray analysis (ZDENEK VALENTA 2012; PARK et al. 2015; MUSE et al. 2017; CHEN et al. 2019; GOBBI et al. 2019).

Therefore, in this current study, DEGs between patients with MI and health control were identified, following by univariable logistic regression, the LASSO, SignalP 3.0 server and multivariable logistic regression. By using ROC curve analyses, area under the curve (AUC) and C-index, a robust MI diagnosis-related gene signature was used to estimate the diagnostic value of genes in patients with MI. Subsequently, the diagnosis-related gene signature was validated in 7 independent validation data sets. Furthermore, the accuracy of diagnosis-related gene signature was further explored for its accuracy to discriminate MI from healthy control by meta-analysis in all data sets.

2. Materials And Methods

2.1 Data mining based on the GEO database

Initially, microarray data were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) up to December 2019. The following search word was used: myocardial infarction. Microarray data were considered eligible if they were case-control that reported the gene expression profiling between patients with MI and healthy control. Exclusion criteria were as follows: (1) duplicate microarray data, (2) lack of case-control, (3) non-human data, (4) the sample of data less than 12 (OBUCHOWSKI and MCCLISH 1997). According to the inclusion criteria, seven GEO datasets were identified and included (see Table 1). Figure 1 described a flow diagram of the GEO datasets selection for this study. For those available datasets, the normalised data of gene expression profiling were downloaded from the database of GEO.

Table 1
Information on the included microarray datasets.

GEO accession	Country	Platform	Cases/controls	Source of tissue
GSE141512	Russia	GPL17586	6/6	Whole Blood
GSE24519	Italy	GPL2895	34/4	Whole Blood
GSE34198	Czech Republic	GPL6102	49/48	Whole Blood
GSE48060	USA	GPL570	31/21	Whole Blood
GSE60993	South Korea	GPL6884	17/7	Whole Blood
GSE66360	USA	GPL570	49/50	CD146 + Circulating Endothelial Cells
GSE109048	Italy	GPL17586	19/19	Platelets

2.2 Identification of a diagnosis-related gene signature set associated with MI

The dataset of GSE66360(MUSE et al. 2017) was the most abundant samples and categorised into the training cohort, which was used to investigate a diagnosis-related gene signature set associated with MI. Initially, the DEGs between MI and control, we used *edgeR* package in R statistical software with the threshold of false rate (FDR) < 0.05 and |log fold change (logFC)|>2. Then, those DEGs with a statistical significance in univariable logistic regression were selected into the least absolute shrinkage and selection operator (LASSO) to obtain first-rank diagnostic genes from the patients with MI. Afterwards, to be clinically detectable serum biomarkers from patients with MI in the future, the optimal diagnostic genes were investigated to identify in the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP-3.0/) (BENDTSEN et al. 2004; EMANUELSSON et al. 2007). Lastly, multivariable logistic regression was utilized to build a diagnosis-related gene signature by incorporating the detectable feature selected from the peripheral blood of patients in the SignalP 3.0 server. Receiver operating characteristic (ROC) curve analyses and area under the curve (AUC) was used to estimate the diagnostic value of a diagnosis-related gene signature in patients with MI and control. Moreover, Harrell's C-index was calculated to quantify the discrimination performance of the diagnosis-related gene signature. *P*-value of < 0.05 was defined to have statistical significance in the analysis.

2.3 Functional and pathway enrichment analysis

Among the DEGs, dataset functional analysis was performed using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses using *ClusterProfiler* and *org.Hs.eg.db* package(YU et al. 2012). GO terms and KEGG pathways with a $p < 0.05$ were considered statistically significant.

2.4 Validation of a diagnosis-related gene signature

Six data sets including GSE141512, GSE24519, GSE34198, GSE48060, GSE60993 and GSE109048 were categorized into the validation sets. To validate whether the candidate genes might have certain important diagnostic value in patients with MI, we also measured ROC, AUC and C-index in the validation sets.

2.5 Meta-analysis

The sensitivity and specificity of each dataset included were calculated by the constructed model of a diagnosis-related gene signature. Then, true positives, false negatives, false positives, and true negatives were tabulated and stratified by the included all datasets in patients with MI and control. Later, a meta-analysis was performed to get pooled the pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), the bivariate summary receiver operator characteristic (SROC) curve, and area under the curve (AUC), which indicated the overall diagnostic value of a diagnosis-related gene signature in distinguishing patients with MI from control. Statistical heterogeneity among the datasets was assessed using by using Cochran's Q statistic and I^2 tests. Values of 25%, 50%, and 75% for the I^2 test were suggestive of the presence of the low, medium, and high significant heterogeneity, respectively. Besides, Fagan's nomogram were used to reveal the clinical utility of the diagnosis-related gene signature. Meta-regression analysis was performed to investigate the effects of potential factors on the diagnostic ability of MI. We assessed the publication bias of the included datasets by using Deeks' regression test of funnel plot asymmetry. All statistical analyses were conducted using STATA 14.0 (Stata Corp, College Station, TX, USA). Meta-DiSc 1.4 (XI Cochrane Colloquium, Barcelona, Spain) were used for threshold effect. A p-value of < 0.05 was considered statistically significant (ZHAO et al. 2019).

3. Results

3.1 Identification of a diagnosis-related gene signature for MI

A total of 44 DEGs were obtained and identified by the genes profiling data of the discovery group to univariable logistic regression analysis (Fig. 2A). Among them, 8 DEGs were selected by the LASSO method for further investigation in the SignalP 3.0 server (Fig. 2B). On the premise of considering signal peptide probability, we identified a total of 3 DEGs, including *CCL20*, *IL1R2* and *ITLN1*. Then, the 3 DEGs were analysed by multivariable logistic regression, which results showed that *CCL20, IL1R2* and *ITLN1* remained significantly associated with MI (Fig. 2D). Next, we used *CCL20*, *IL1R2* and *ITLN1* to construct a diagnosis-related gene signature for distinguishing patients with MI from healthy control. Considering the discrimination ability of a diagnosis-related gene signature, ROC analysis was conducted. The results showed that the sensitivity, specificity and AUC were 0.918, 0.980 and 0.975 to suggest a diagnosis-related gene signature for MI with higher prediction efficacy. What's more, the C-index value of 0.975 for the 3 DEGs in patients with MI also indicated good discrimination.

3.2 Validation of the three-gene signature in six independent cohorts

The robustness of the three-gene signature was regarded as a candidate biomarker for predicting diagnosis in patients with MI, while the validation cohort consisted of the remaining GSE141512, GSE24519, GSE34198, GSE48060, GSE60993 and GSE109048 data sets. However, the results of AUC for the validation cohort showed that the three-gene signature was differently predictive power. Four data sets showed good accuracy in predicting MI (AUC = 0.78 in GSE48060, AUC = 0.978 in GSE24519, AUC = 0.882 in GSE60993 and AUC = 0.867 in GSE109048), but the rest of data sets had a weak predictive power (AUC = 0.639 in GSE141512 and AUC = 0.652 in GSE34198). The results of the sensitivity, specificity for the validation cohort were also displayed in Table 2, which also indicated that the ability of the three-gene signature to distinguish MI from control was the same as the manifestation of AUC. Also, the results of C-index for the six data sets were similar to the effects of AUC for them (Table 2).

Table 2
Sensitivity, specificity, AUC and C-index of the classification performance of the three-gene signature in six datasets.

GEO accession	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	C-index (95% CI)
GSE141512	3	1	3	5	0.500 (0.139–0.860)	0.833 (0.364–0.991)	0.639 (0.311–0.967)	0.639 (0.00788-1.00)
GSE24519	32	0	2	4	0.941 (0.789–0.989)	1.00 (0.395-1.000)	0.978 (0.934-1.00)	0.978 (0.899-1.00)
GSE34198	27	10	22	38	0.551 (0.403–0.691)	0.79 (0.645–0.890)	0.652 (0.542–0.762)	0.652 (0.446–0.857)
GSE48060	24	4	7	17	0.774 (0.584–0.897)	0.809 (0.574–0.937)	0.78 (0.641–0.920)	0.78 (0.501-1.00)
GSE60993	15	1	2	6	0.824 (0.558–0.953)	1.00 (0.561-1.00)	0.882 (0.744-1.000)	0.882 (0.613-1.00)
GSE66360	45	1	4	49	0.918 (0.795–0.973)	0.980 (0.879–0.998)	0.975 (0.948-1.000)	0.975 (0.922-1.00)
GSE109048	14	2	5	17	0.736 (0.485–0.898)	0.894 (0.654–0.981)	0.867 (0.749–0.985)	0.867 (0.635-1.00)

3.3 Functional annotation

Analysis of the three-gene signature by GO categories and KEGG pathways was crucial for our understanding of biological function. In this study, the top enriched GO terms for biological process (BP) were as follows: cellular response to interleukin-1, response to interleukin-1 and negative regulation of

interleukin-1 secretion; and for molecular function (MF): RAGE receptor binding, Toll-like receptor binding and carbohydrate-binding (Table 3). Functional enrichment analysis showed that the top 20 KEGG pathways included the chemokine signaling pathway, IL − 17 signaling pathway and TNF signaling pathway (Table 4).

Table 3
GO functional annotation of the three-gene signature.

Category	ID	GO term	*p*-value	Gene
BP	GO:0071347	cellular response to interleukin-1	0.00014	*CCL20, IL1R2*
BP	GO:0070555	response to interleukin-1	0.00020	*CCL20, IL1R2*
BP	GO:0050711	negative regulation of interleukin-1 secretion	0.0025	*IL1R2*
BP	GO:1900016	negative regulation of cytokine production involved in inflammatory response	0.0031	*IL1R2*
BP	GO:0035584	calcium-mediated signaling using intracellular calcium source	0.0035	*CCL20*
BP	GO:0032692	negative regulation of interleukin-1 production	0.0045	*IL1R2*
BP	GO:2000406	positive regulation of T cell migration	0.0048	*CCL20*
BP	GO:0070207	protein homotrimerization	0.0052	*ITLN1*
BP	GO:0046326	positive regulation of glucose import	0.0053	*ITLN1*
BP	GO:2000403	positive regulation of lymphocyte migration	0.0058	*CCL20*
BP	GO:0010955	negative regulation of protein processing	0.0061	*IL1R2*
BP	GO:1903318	negative regulation of protein maturation	0.0061	*IL1R2*
BP	GO:0010828	positive regulation of glucose transmembrane transport	0.0064	*ITLN1*
BP	GO:1900015	regulation of cytokine production involved in inflammatory response	0.0064	*IL1R2*
BP	GO:2000404	regulation of T cell migration	0.0064	*CCL20*
BP	GO:0002534	cytokine production involved in inflammatory response	0.0069	*IL1R2*
BP	GO:0050704	regulation of interleukin-1 secretion	0.0079	*IL1R2*
BP	GO:0070206	protein trimerization	0.0087	*ITLN1*
BP	GO:0046324	regulation of glucose import	0.0088	*ITLN1*
BP	GO:0070498	interleukin-1-mediated signaling pathway	0.0088	*IL1R2*
BP	GO:0001960	negative regulation of cytokine-mediated signaling pathway	0.0090	*IL1R2*
BP	GO:0050701	interleukin-1 secretion	0.0090	*IL1R2*
BP	GO:0072678	T cell migration	0.0095	*CCL20*

Category	ID	GO term	*p*-value	Gene
BP	GO:2000401	regulation of lymphocyte migration	0.0095	*CCL20*
BP	GO:0060761	negative regulation of response to cytokine stimulus	0.0097	*IL1R2*
BP	GO:0002532	production of molecular mediator involved in inflammatory response	0.010	*IL1R2*
BP	GO:0046323	glucose import	0.010	*ITLN1*
BP	GO:0048247	lymphocyte chemotaxis	0.010	*CCL20*
BP	GO:0002548	monocyte chemotaxis	0.010	*CCL20*
BP	GO:0050710	negative regulation of cytokine secretion	0.010	*IL1R2*
BP	GO:0010827	regulation of glucose transmembrane transport	0.011	*ITLN1*
BP	GO:0032652	regulation of interleukin-1 production	0.012	*IL1R2*
BP	GO:0071674	mononuclear cell migration	0.013	*CCL20*
BP	GO:0032612	interleukin-1 production	0.014	*IL1R2*
BP	GO:0070098	chemokine-mediated signaling pathway	0.014	*CCL20*
BP	GO:1990868	response to chemokine	0.015	*CCL20*
BP	GO:1990869	cellular response to chemokine	0.015	*CCL20*
BP	GO:0030593	neutrophil chemotaxis	0.016	*CCL20*
BP	GO:1904659	glucose transmembrane transport	0.016	*ITLN1*
BP	GO:0072676	lymphocyte migration	0.016	*CCL20*
BP	GO:0008645	hexose transmembrane transport	0.017	*ITLN1*
BP	GO:0015749	monosaccharide transmembrane transport	0.017	*ITLN1*
BP	GO:0034219	carbohydrate transmembrane transport	0.018	*ITLN1*
BP	GO:1990266	neutrophil migration	0.018	*CCL20*
BP	GO:0019730	antimicrobial humoral response	0.019	*ITLN1*
BP	GO:0071621	granulocyte chemotaxis	0.019	*CCL20*
BP	GO:0002687	positive regulation of leukocyte migration	0.019	*CCL20*
BP	GO:0050709	negative regulation of protein secretion	0.020	*IL1R2*
BP	GO:0097530	granulocyte migration	0.021	*CCL20*
BP	GO:0002792	negative regulation of peptide secretion	0.021	*IL1R2*

Category	ID	GO term	*p*-value	Gene
BP	GO:0050728	negative regulation of inflammatory response	0.022	*IL1R2*
BP	GO:0008643	carbohydrate transport	0.022	*ITLN1*
BP	GO:0001959	regulation of cytokine-mediated signaling pathway	0.024	*IL1R2*
BP	GO:0060759	regulation of response to cytokine stimulus	0.026	*IL1R2*
BP	GO:0070613	regulation of protein processing	0.027	*IL1R2*
BP	GO:1903317	regulation of protein maturation	0.027	*IL1R2*
BP	GO:0071346	cellular response to interferon-gamma	0.028	*CCL20*
BP	GO:0002685	regulation of leukocyte migration	0.028	*CCL20*
BP	GO:0051224	negative regulation of protein transport	0.029	*IL1R2*
BP	GO:0050707	regulation of cytokine secretion	0.029	*IL1R2*
BP	GO:1904950	negative regulation of establishment of protein localization	0.029	*IL1R2*
BP	GO:0034764	positive regulation of transmembrane transport	0.030	*ITLN1*
BP	GO:1903531	negative regulation of secretion by cell	0.031	*IL1R2*
BP	GO:0034341	response to interferon-gamma	0.031	*CCL20*
BP	GO:0031348	negative regulation of defense response	0.031	*IL1R2*
BP	GO:0097529	myeloid leukocyte migration	0.031	*CCL20*
BP	GO:0050663	cytokine secretion	0.033	*IL1R2*
BP	GO:0030595	leukocyte chemotaxis	0.034	*CCL20*
BP	GO:0019722	calcium-mediated signaling	0.035	*CCL20*
BP	GO:0051048	negative regulation of secretion	0.035	*IL1R2*
BP	GO:0071356	cellular response to tumor necrosis factor	0.037	*CCL20*
MF	GO:0070492	oligosaccharide binding	0.0025	*ITLN1*
MF	GO:0048020	CCR chemokine receptor binding	0.0072	*CCL20*
MF	GO:0008009	chemokine activity	0.0083	*CCL20*
MF	GO:0042379	chemokine receptor binding	0.011	*CCL20*
MF	GO:0004896	cytokine receptor activity	0.015	*IL1R2*
MF	GO:0019955	cytokine binding	0.021	*IL1R2*

Category	ID	GO term	*p*-value	Gene
MF	GO:0019838	growth factor binding	0.023	*IL1R2*

Table 4
KEGG pathway analysis of the three-gene signature.

ID	KEGG term	*p*-value	Gene
hsa04060	Cytokine-cytokine receptor interaction	0.0010	*CCL20*, *IL1R2*
hsa05323	Rheumatoid arthritis	0.023	*CCL20*
hsa04657	IL-17 signaling pathway	0.023	*CCL20*
hsa05215	Prostate cancer	0.024	*IL1R2*
hsa04640	Hematopoietic cell lineage	0.024	*IL1R2*
hsa04061	Viral protein interaction with cytokine and cytokine receptor	0.025	*CCL20*
hsa05146	Amoebiasis	0.025	*IL1R2*
hsa04668	TNF signaling pathway	0.028	*CCL20*
hsa05418	Fluid shear stress and atherosclerosis	0.034	*IL1R2*
hsa04062	Chemokine signaling pathway	0.046	*CCL20*
hsa05202	Transcriptional misregulation in cancer	0.047	*IL1R2*
hsa05166	Human T-cell leukemia virus 1 infection	0.054	*IL1R2*

3.4 Meta-analysis for diagnosis

A total of 7 data sets were included in the meta-analysis to determine the diagnostic value of the three-gene signature. As shown in Fig. 4, the pooled sensitivity and specificity estimates for the three-gene signature were 0.80 (95% CI: 0.66–0.90) and 0.90 (95% CI: 0.80–0.96), respectively. The moderate informational value of the three-gene signature implied a PLR (8.4), but the NLR (0.22) indicated minimal informational value. Figure 4D displays the use of the likelihood ratio scattergram for investigating diagnostic value; when the right lower quadrant was depicted, the three-gene signature was useful for confirming the presence of MI (while positive) but not for its exclusion (while negative). The DOR and area under the ROC curve were 39 (95% CI: 9-159) and 0.93 (95% CI: 0.90–0.95), respectively, which indicated that the three-gene signature has good discriminatory ability for MI. Figure 4C depicts the use of Fagan's nomogram for calculating posttest probabilities; the three-gene signature increased the likelihood of MI from 57–92%, and the risk decreased to 22% when a negative result was confirmed.

Significant heterogeneity was observed (81.54% for sensitivity and 58.99% for specificity) among the 7 included data sets. Thus, to identify the source of heterogeneity, we analyzed heterogeneity from the aspect of a threshold effect, publication bias, bivariate box plot, and meta-regression. The Spearman correlation analysis (correlation coefficient =-0.714, $p = 0.071$) revealed no threshold effect on the three-gene signature for distinguishing patients with MI from healthy control. Deeks' funnel plot asymmetry test demonstrated no potential publication bias in our data sets (t = -0.30; p-value = 0.77) (Fig. 5A). The bivariate box plot revealed that the central location included 6 data sets, with one data set as the outlier, suggesting a low degree of indirect heterogeneity (Fig. 5B). Then, meta-regression was performed to analyze patient size, location, source of the tissue, median distribution and platforms. The major sources of heterogeneity for specificity were the source of the source of the tissue and median distribution. However, the potential sources of heterogeneity for sensitivity were not confirmed. The metaregression results are shown in Fig. 6.

4. Discussion

In the present study, it was found that the higher expression of *CCL20*, *IL1R2* and *ITLN1* in patients with MI compared with healthy control were utilized to construct the model, which had an excellent diagnostic performance for patients in the 7 data sets. The additional diagnostic meta-analysis demonstrated that the three-gene signature turned an outstanding performance in predicting the diagnosis of MI patients. In this research, the ROC area of the three-gene signature was 0.93, which indicated that the three-gene signature might be considered as the candidate therapeutic targets for MI patients. Interestingly, CCL20, IL1R2 and ITLN1 possessed the characteristic of the secretory molecule by using the SignalP 3.0 analysis. So the highly expressed of CCL20, IL1R2 and ITLN1 might be measured in the blood and provided to the early diagnostic biomarkers for MI.

Recently, an increasing number of studies have previously shown that CCL20, IL1R2 and ITLN1 were found to be correlated with MI(YAN et al. 2012; KATAOKA et al. 2014; KADOGLOU et al. 2015; MENZEL et al. 2016; SAFA et al. 2016; STEJSKAL et al. 2016; YABLUCHANSKIY et al. 2016; NAGASAKA et al. 2017; CHANG et al. 2018; LIAN et al. 2018; ORREM et al. 2018; CHEN et al. 2019; LIN et al. 2019; ZHU et al. 2019; ZHANG et al. 2020). It was uncovered that the stimulation by transforming growth factor-β of peripheral blood mononuclear cells could have enhanced expression of CCL20 (LIN et al. 2019). Additionally, the trend of increased serum levels of CCL20 in patients with MI was not significantly increased compared to healthy control. A tremendous potential reason for its result might be not big enough size of the sample used to conduct statistical significance. However, one previous study demonstrated that serum levels of CCL20 were significantly higher in patients with ischemic heart disease, which included acute MI, stable angina and unstable angina (SAFA et al. 2016). What's more, the previous study implied that T-cell death-associated gene 8 (TDAG8) negatively regulated the transcription of the chemokine Ccl20, subsequently increasing expression of CCL20 in TDAG8 KO mice and eventually contributing to the survival rate and cardiac function by suppressing CCL20 (Nagasaka *et al.* 2017). It should be noted that the appearance of CCL20 increased after the activated mitogen-activated protein kinase by the stimulation of IL-17 signalling. When CCL20 bound to the CCR6 receptor, it played an

essential role to recruit the chemoattraction of leukocyte and mediate γδT cells to the inflammation locus and thus the aggravation of cardiac function(YAN et al. 2012; CHANG et al. 2018).

IL1R2 is involved in the process of coronary atherosclerosis. For example, Lian *et al*. reported that IL1R2 was mediated by miR-383-3p to prevent inflammation injury in the inflammatory damage of coronary artery endothelial cells through the inhibition of the activation of an inflammasome signalling pathway (LIAN et al. 2018). IL1R2 has two different forms of the protein including membrane-binding protein and soluble-protein (soluble IL-1 receptor 2), which was significantly associated with the left ventricular remodelling in patients with ST-elevation MI (ORREM et al. 2018).

Omentin-1, also referred to ITLN1, was a novel adipokine which was related to the processes of glucose metabolism, inflammation, and atherosclerosis (MENZEL et al. 2016). Circulating omentin is associated with coronary artery disease (SHIBATA et al. 2011). Shibata et al. indicated that low levels of ITLN1 were related to coronary artery disease (CAD) and that ITLN1can be considered as a novel biomarker for CAD (SHIBATA et al. 2011). Similarly, Kadoglou *et al*. identified that ITLN1 appeared with significantly lower in patients with acute MI at admission, but significantly higher through the suppression of inflammation after six months at the hospital, which implied ITLN1may be as a novel target of treatment (KADOGLOU et al. 2015). On the contrary, Menzel *et al*. shown that ITLN1was not significantly associated with risk of MI after multivariable adjustment (MENZEL et al. 2016).

There are several shortcomings, which should be considered in our study. The main limitation of this study was conducted using data with the small sample size, which came from several published data sets. Our findings need to validate in other data sets and clinical trials whether CCL20, IL1R2 and ITLN1 may provide on biomarkers for MI. Moreover, the three-gene signature was based on only in silico methods and only a fraction of the human gene was included in the analysis. So the diagnostic genes could remain not to represent all the gene candidates that were potentially associated with MI. Finally, the mechanisms through which the three-gene signature modulate the progression of MI were necessary to make further investigation. Despite these drawbacks, however, this study still provides a potentially powerful diagnostic marker for MI.

5. Conclusions

In summary, by combining CCL20, IL1R2 and ITLN1, the three-gene signature was significantly associated with diagnosis in MI and could provide potential therapeutic targets and novel therapeutic strategies for MI.

Abbreviations

MI
Myocardial infarction; cTnT:cardiac troponin T; CK-MB:creatine kinase MB; GEO:Gene Expression Omnibus;DEGs:differentially expressed genes; ROC:Receiver operating characteristic; AUC:area under the

curve; LASSO:least absolute shrinkage and selection operator; GO:gene ontology; KEGG:Kyoto Encyclopedia of Genes and Genomes; PLR:positive likelihood ratio; NLR:negative likelihood ratio; DOR:diagnostic odds ratio; CAD:coronary artery disease;

Declarations

Ethical Approval and Consent to participate

There was no ethical approval and consent to participate.

Consent for publication

All authors of this paper have read and approved the final version submitted.

Availability of supporting data

The data that support the findings of this study are made available in the Gene Expression Omnibus (GEO) database (GSE66360, GSE141512, GSE24519, GSE34198, GSE48060, GSE60993 and GSE109048).

Competing interests

All authors declare that they have no conflict of interest.

Funding

This project was sponsored by the Key Project of Education Department of Hebei Province (Grant no. ZD20161011), the Hebei Provincial Natural Science Foundation (Grant no. H2018406061), and the Key Project of Health Commission of Hebei Province (Grant no. 20181153).

Authors' contributions

Research conception and design: Jingyi Zhao; Data analysis and interpretation: Xiaohui Zhou and Ying Wang; Drafting of the manuscript: Yinhui Yao; Critical revision of the manuscript: Jingyi Zhao; Approval of final manuscript: all authors.

Acknowledgements

Not applicable.

Authors' information

Yinhui Yao, Department of Pharmacy, the Affiliated Hospital of Chengde Medical College.Email: yaoyh_gc@163.com.

Jingyi Zhao, Department of Functional Center, Chengde Medical College. Email: 33286986@qq.com.

Xiaohui Zhou, School of Basic Medicine, Chengde Medical College. Email: 2459220834@qq.com.

Ying Wang, Department of Pharmacy, the Affiliated Hospital of Chengde Medical College. Email: zgzz0314@sohu.com.

References

BENDTSEN, J. D., H. NIELSEN, G. VON HEIJNE and S. BRUNAK, 2004 Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783-795.

CHAN, D., and L. L. NG, 2010 Biomarkers in acute myocardial infarction. BMC Med 8: 34.

CHANG, S. L., Y. W. HSIAO, Y. N. TSAI, S. F. LIN, S. H. LIU et al., 2018 Interleukin-17 enhances cardiac ventricular remodeling via activating MAPK pathway in ischemic heart failure. J Mol Cell Cardiol 122: 69-79.

CHEN, D. Q., X. S. KONG, X. B. SHEN, M. Z. HUANG, J. P. ZHENG et al., 2019 Identification of Differentially Expressed Genes and Signaling Pathways in Acute Myocardial Infarction Based on Integrated Bioinformatics Analysis. Cardiovasc Ther 2019: 8490707.

CHRISTENSON, R. H., R. T. VOLLMER, E. M. OHMAN, S. PECK, T. D. THOMPSON et al., 2000 Relation of temporal creatine kinase-MB release and outcome after thrombolytic therapy for acute myocardial infarction. TAMI Study Group. Am J Cardiol 85: 543-547.

DE WINTER, R. J., R. W. KOSTER, A. STURK and G. T. SANDERS, 1995 Value of myoglobin, troponin T, and CK-MBmass in ruling out an acute myocardial infarction in the emergency room. Circulation 92: 3401-3407.

EMANUELSSON, O., S. BRUNAK, G. VON HEIJNE and H. NIELSEN, 2007 Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2: 953-971.

GOBBI, G., C. CARUBBI, G. M. TAGLIAZUCCHI, E. MASSELLI, P. MIRANDOLA et al., 2019 Sighting acute myocardial infarction through platelet gene expression. Sci Rep 9: 19574.

KADOGLOU, N. P., D. K. TAHMATZIDIS, C. GIANNAKOULAS, A. KAPELOUZOU, A. GKONTOPOULOS et al., 2015 Serum levels of novel adipokines, omentin-1 and chemerin, in patients with acute myocardial infarction: KOZANI STUDY. J Cardiovasc Med (Hagerstown) 16: 341-346.

KATAOKA, Y., R. SHIBATA, K. OHASHI, T. KAMBARA, T. ENOMOTO et al., 2014 Omentin prevents myocardial ischemic injury through AMP-activated protein kinase- and Akt-dependent mechanisms. J Am Coll Cardiol 63: 2722-2733.

LIAN, Z., F. F. LV, J. YU and J. W. WANG, 2018 The anti-inflammatory effect of microRNA-383-3p interacting with IL1R2 against homocysteine-induced endothelial injury in rat coronary arteries. J Cell Biochem 119: 6684-6694.

LIN, C. F., C. J. SU, J. H. LIU, S. T. CHEN, H. L. HUANG et al., 2019 Potential Effects of CXCL9 and CCL20 on Cardiac Fibrosis in Patients with Myocardial Infarction and Isoproterenol-Treated Rats. J Clin Med 8.

LIU, N., F. ZHENG and X. ZHENG, 2018 Detection of biomarkers of acute myocardial infarction by high-throughput suspension array technology in serum sample. Bioanalysis 10: 47-58.

MENZEL, J., R. DI GIUSEPPE, R. BIEMANN, C. WITTENBECHER, K. ALEKSANDROVA et al., 2016 Omentin-1 and risk of myocardial infarction and stroke: Results from the EPIC-Potsdam cohort study. Atherosclerosis 251: 415-421.

MUSE, E. D., E. R. KRAMER, H. WANG, P. BARRETT, F. PARVIZ et al., 2017 A Whole Blood Molecular Signature for Acute Myocardial Infarction. Sci Rep 7: 12268.

NAGASAKA, A., C. MOGI, H. ONO, T. NISHI, Y. HORII et al., 2017 The proton-sensing G protein-coupled receptor T-cell death-associated gene 8 (TDAG8) shows cardioprotective effects against myocardial infarction. Sci Rep 7: 7812.

OBUCHOWSKI, N. A., and D. K. MCCLISH, 1997 Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. Stat Med 16: 1529-1542.

ORREM, H. L., C. SHETELIG, T. UELAND, S. LIMALANATHAN, P. H. NILSSON et al., 2018 Soluble IL-1 receptor 2 is associated with left ventricular remodelling in patients with ST-elevation myocardial infarction. Int J Cardiol 268: 187-192.

PARK, H. J., J. H. NOH, J. W. EUN, Y. S. KOH, S. M. SEO et al., 2015 Assessment and diagnostic relevance of novel serum biomarkers for early decision of ST-elevation myocardial infarction. Oncotarget 6: 12970-12983.

RAKOWSKI, T., A. DZIEWIERZ, J. LEGUTKO, P. KLECZYNSKI, A. BRZOZOWSKA-CZARNEK et al., 2014 Creatine kinase-MB assessed in patients with acute myocardial infarction correlates with cardiac magnetic resonance infarct size at 6-month follow up. Hellenic J Cardiol 55: 4-8.

SAFA, A., H. R. RASHIDINEJAD, M. KHALILI, S. DABIRI, M. NEMATI et al., 2016 Higher circulating levels of chemokines CXCL10, CCL20 and CCL22 in patients with ischemic heart disease. Cytokine 83: 147-157.

SHIBATA, R., N. OUCHI, R. KIKUCHI, R. TAKAHASHI, K. TAKESHITA et al., 2011 Circulating omentin is associated with coronary artery disease in men. Atherosclerosis 219: 811-814.

STEJSKAL, D., J. VACLAVIK, A. SMEKAL, G. SVOBODOVA, R. RICHTEROVA et al., 2016 Omentin-1 levels in patients with premature coronary artery disease, metabolic syndrome and healthy controls. Short communication. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 160: 219-221.

WANG, C., and Q. JING, 2018 Non-coding RNAs as biomarkers for acute myocardial infarction. Acta Pharmacol Sin 39: 1110-1119.

YABLUCHANSKIY, A., Y. MA, K. Y. DELEON-PENNELL, R. ALTARA, G. V. HALADE et al., 2016 Myocardial Infarction Superimposed on Aging: MMP-9 Deletion Promotes M2 Macrophage Polarization. J Gerontol A

Biol Sci Med Sci 71: 475-483.

YAN, X., T. SHICHITA, Y. KATSUMATA, T. MATSUHASHI, H. ITO et al., 2012 Deleterious effect of the IL-23/IL-17A axis and γδT cells on left ventricular remodeling after myocardial infarction. J Am Heart Assoc 1: e004408.

YU, G., L. G. WANG, Y. HAN and Q. Y. HE, 2012 clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16: 284-287.

ZDENEK VALENTA, I. M., MICHAL KOLAR, HANA GRUNFELDOVA, PETRA FEGLAROVA, J PELESKA, M TOMECKOVA, JAN KALINA, DALIBOR SLOVAK, J ZVAROVA, 2012 Determinants of excess genetic risk of acute myocardial infarction-a matched case-control study. European Journal for Biomedical Informatics 8: 34-43.

ZHANG, R., Z. JI, Y. QU, M. YANG, Y. SU et al., 2020 Clinical value of ARG1 in acute myocardial infarction patients: Bioinformatics-based approach. Biomed Pharmacother 121: 109590.

ZHAO, J., H. YU, P. YAN, X. ZHOU, Y. WANG et al., 2019 Circulating MicroRNA-499 as a Diagnostic Biomarker for Acute Myocardial Infarction: A Meta-analysis. Dis Markers 2019: 6121696.

ZHU, Y., C. HU, Y. DU, J. ZHANG, J. LIU et al., 2019 Time-Dependent Change in Omentin-1 Level Correlated with Early Improvement of Myocardial Function in Patients with First Anterior ST-Segment Elevation Myocardial Infarction After Primary Percutaneous Coronary Intervention. J Atheroscler Thromb 26: 856-867.

Figures

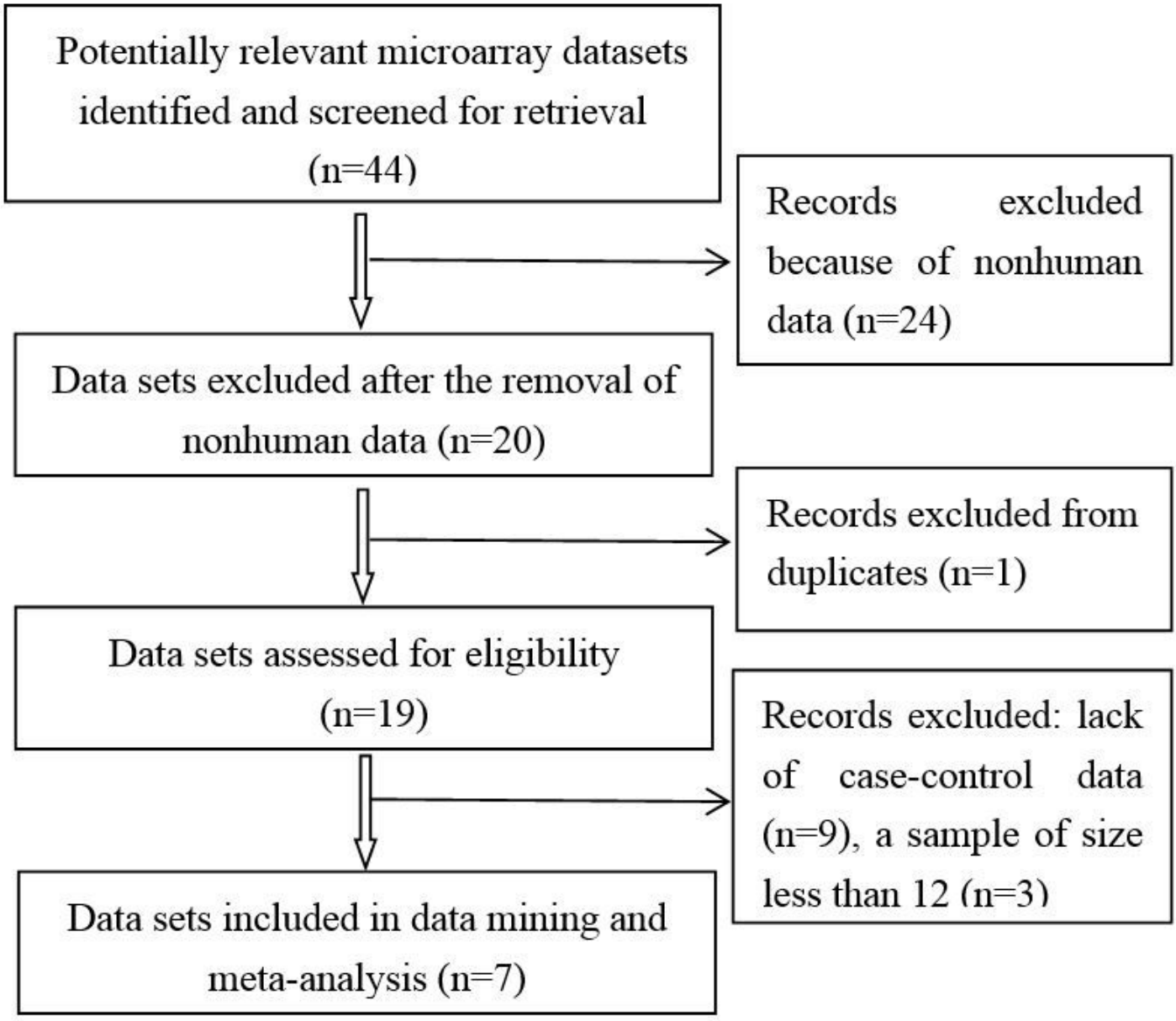

Figure 1

Flow chart of microarray data sets selection.

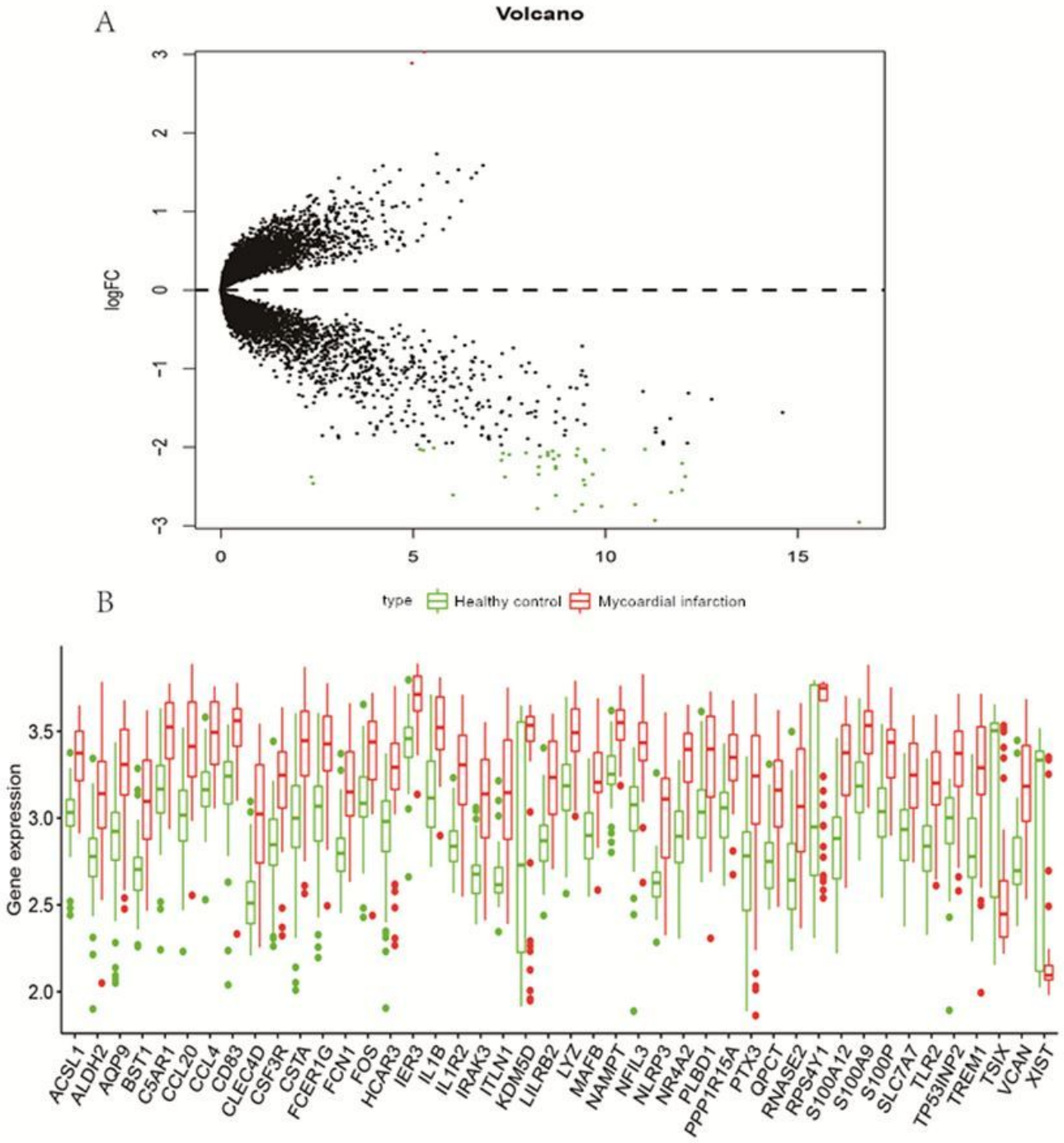

Figure 2

Differentially expressed genes (DEGs) between myocardial infarction (MI) and healthy control tissues. (A) Volcano plot for the 44 DEGs (MI vs healthy control). Red indicates high expression, and green indicates low expression. (B) Expression patterns of 44 DEGs between MI and healthy control.

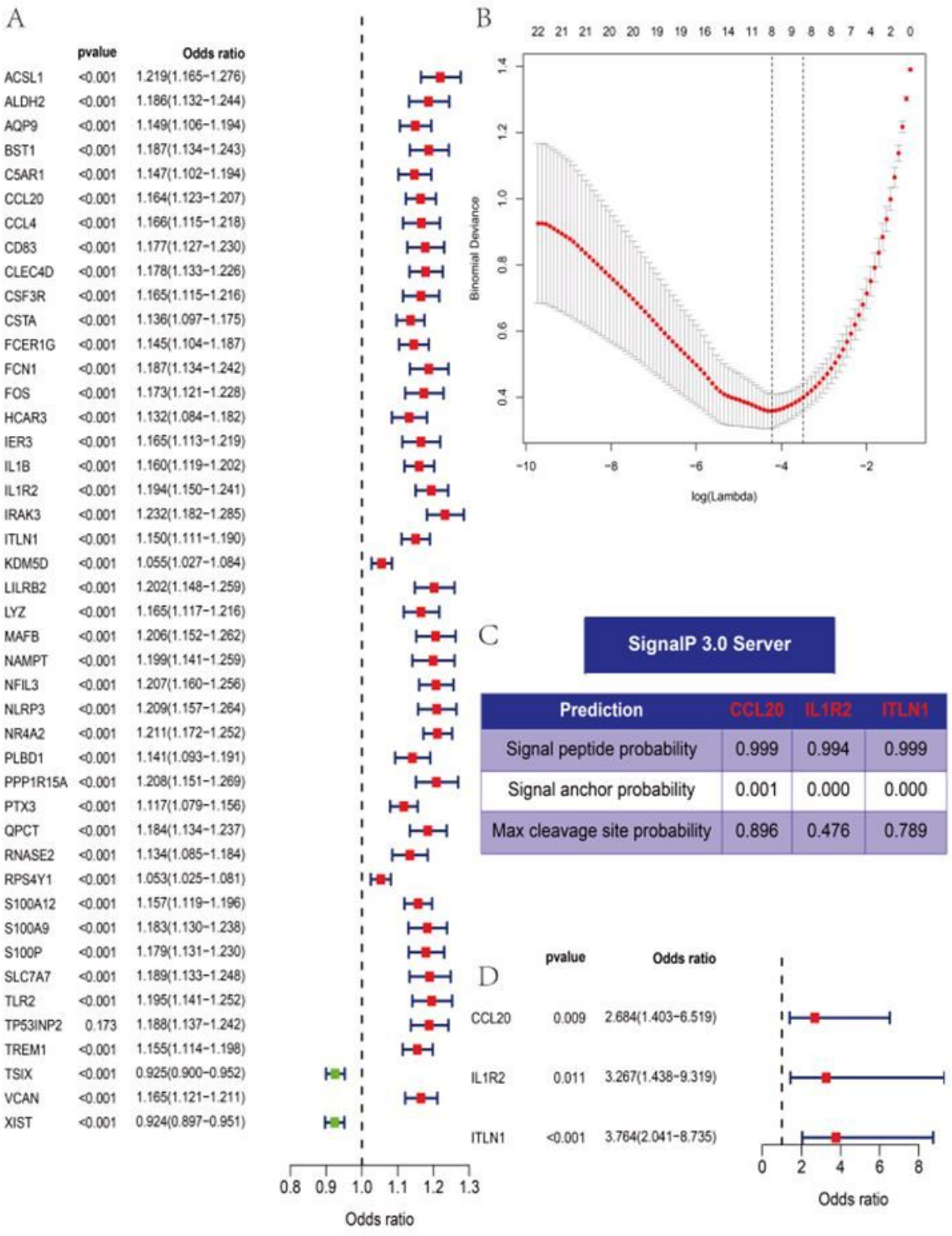

Figure 3

Identification of the three-gene signature for patients with myocardial infarction (MI) in the GSE66360 data set. (A) Forty-four DEGs were analyzed by univariable logistic regression. (B) Eight DEGs were identified by least absolute shrinkage and selection operator (LASSO) regression. (C) Three DEGs were identified by using the SignalP 3.0 server. (D) The three-gene signature was identified by multivariable logistic regression.

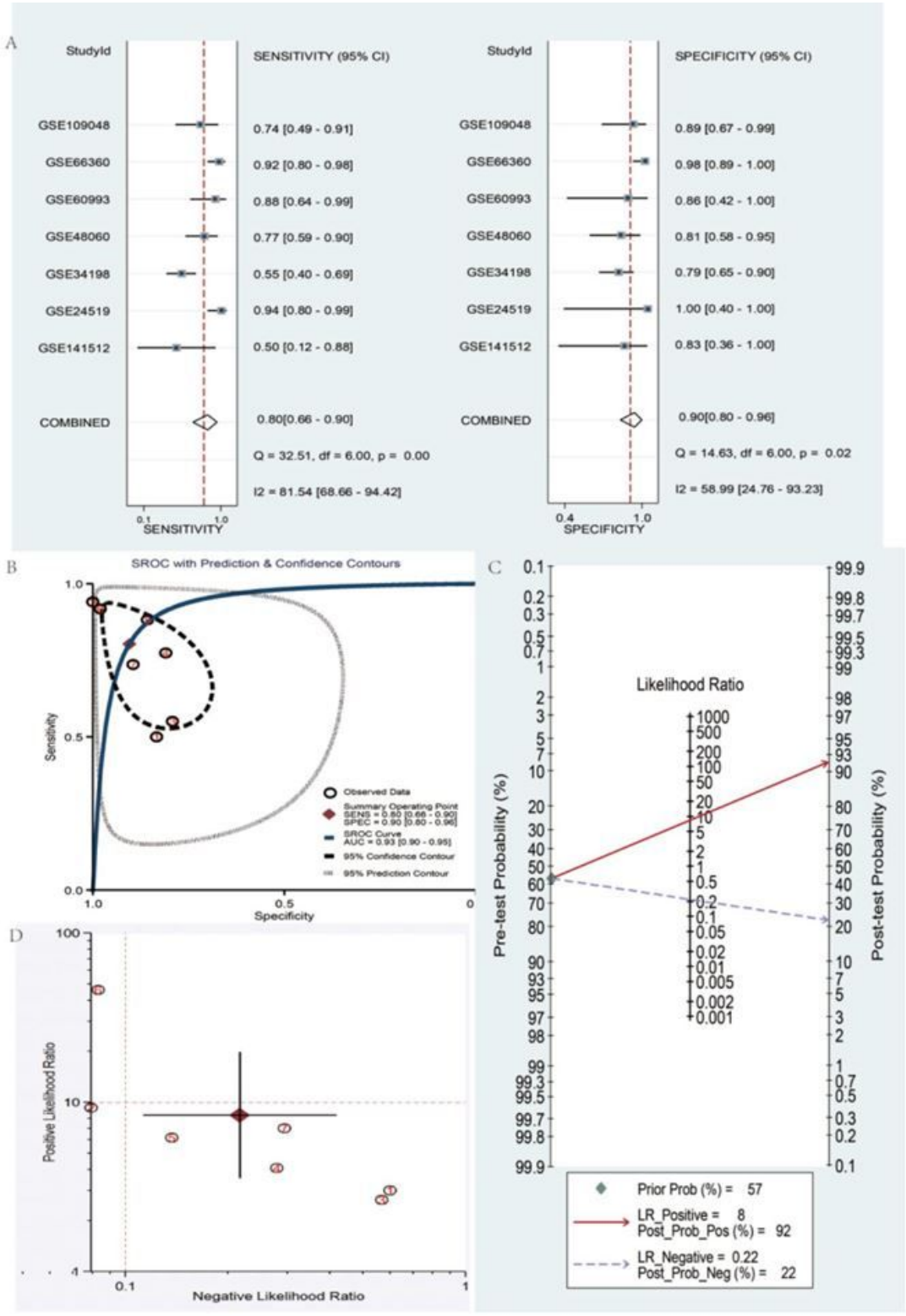

Figure 4

Meta-analysis of the three-gene signature for predicting diagnosis in patients with myocardial infarction (MI). (A) Forest plots of the pooled sensitivity and specificity of the three-gene signature in the diagnosis of MI. (B) Summary receiver operating characteristic (SROC) curve of the three-gene signature. (C) Fagan's nomogram was used to evaluate the clinical utility of the three-gene signature for the diagnosis of MI. (D) Likelihood ratio scattergram.

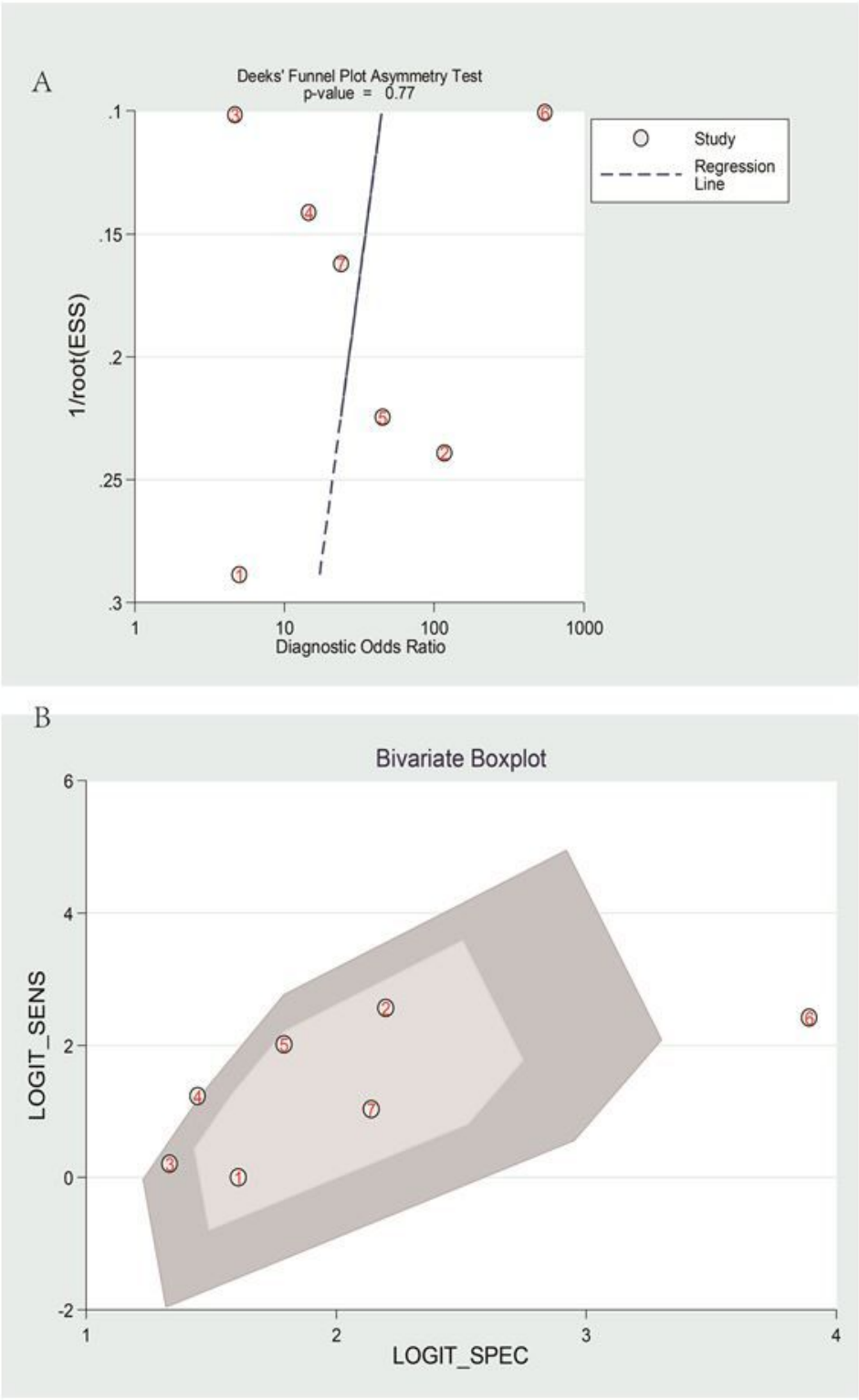

Figure 5

The source of heterogeneity was analyzed from the perspectives of publication bias, and bivariate box plot. (A) Deeks' funnel plot asymmetry test for identifying publication bias. (B) Bivariate boxplot, with most studies clustering within the median distribution with 1 outlier, indirectly suggesting a low degree of heterogeneity.

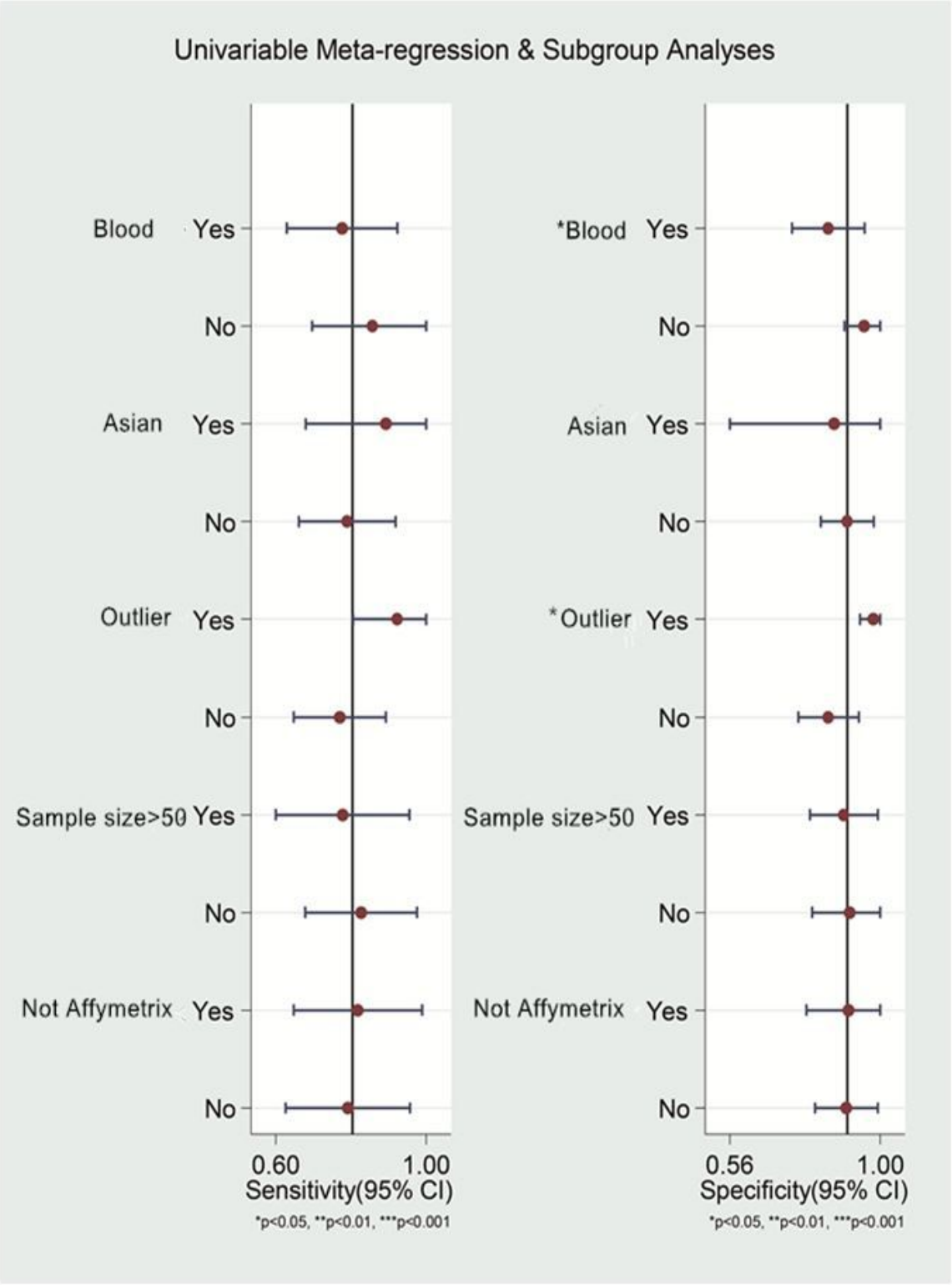

Figure 6

Univariable meta-regression and subgroup analysis in the meta-analysis.