functions in patients following CABG, partially through regulating the pathway of chemokine receptors bind chemokines.

Keywords: coronary artery bypass graft; sevoflurane; module; pathway enrichment analysis

1 Introduction

In recent years, off-pump coronary artery bypass grafting (CABG) surgery has emerged as a common method in treating cardiac disease [1]. However, this less invasive technique can result in the myocardial injury because of a transient period of local ischemia with temporary occlusion of the target vessel, especially in patients with poor cardiac contractile function [2,3]. Thus, improving the perioperative management by reducing myocardial ischemia–reperfusion damage is of paramount importance.

Recent work from clinical trials provides strong evidence that certain anesthetic agents have the potential to decrease myocardial ischemic damage, albeit to a certain degree [4,5]. Sevoflurane, as a kind of inhalational anaesthetic agent, has been indicated to exert a protective function in the preconditioning of cardiac events in patients treated by CABG [6]. Another study has demonstrated that sevoflurane remarkably reduces the size of infarcts as well as Ca$^{2+}$ loading to protect the myocardium against reperfusion injury [5]. In addition, a former study has indicated that sevoflurane provides the myocardial protection in CABG surgery via down-regulating the expression of troponin I [7]. Nevertheless, the underlying molecular mechanisms of sevoflurane for cardioprotection remain to be elucidated.

Currently, microarray analysis is broadly employed to elucidate the underlying mechanisms of various diseases, and is helpful for determining the potential biomarkers of diseases, because of the lower expense and advancements in this technique [8]. Thus, microarray technology might
be useful to determine the effects of sevoflurane on gene expression in patients treated by CABG. In an attempt to determine the influence of sevoflurane on postoperative recovery in patients undergoing CABG, a novel method named as EgoNet [9] on the basis of egocentric network-analysis technique was developed in this study. This method searches significant sub-networks which are functionally related with diseases. The sub-networks identified by our method are defined as modules, which are well-defined in the study of social networks [10]. Firstly, a dataset from the EMBL-EBI database (E-GEOD-4386) was downloaded for subsequent analysis. Then, the differential co-expression networks (DCNs) were constructed. Next, snowball sampling was used to collect all modules. Thereafter, the significance of the identified candidate modules was assessed using random permutation test. Finally, functional analysis for differential modules was carried out to illuminate the biological processes.

## Materials and methods

### 2.1 Gene expression profile

The gene expression data under the series number of E-GEOD-4386 [11] were recruited from the EMBL-EBI database from the platform of A-AFFY-44 -Affymetrix GeneChip Human Genome U133 Plus 2.0 [HG-U133_Plus_2]. There were 40 samples in the data profile E-GEOD-4386, including data from patients treated by CABG plus propofol treatment (n = 10), sevoflurane treatment (n = 10) and control samples (n = 20). The control samples were composed of the same patients before CABG surgery. In the present analysis, to evaluate the influence of sevoflurane on postoperative recovery in patients undergoing CABG, we only selected atrial samples prior to and following CABG with sevoflurane treatment. The average age of patients was 65.2 years. All the patients were male. Moreover, patients with hemodynamic instability were not included. We implemented microarray analysis according to E-GEOD-4386, where total RNA was extracted from the frozen cardiac tissue using an RNeasy Mini kit.

### 2.2 Pretreatment of raw data

The annotation file in TXT format was utilized to map the association between probes and gene symbols. Then, background adjustment was performed using RMA, and quantile normalization was carried out through quantiles. Simultaneously, PM/MM correction was implemented via MAS, followed by summarization of the expression measures using median polish. Next, the probes were mapped to gene symbols, and finally, we obtained 20545 genes.

### 2.3 Protein-protein interaction network (PPIN)

Since complicated biological processes were usually regulated by tightly connected proteins, and the PPIN was further analyzed based on the public database STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, version 9.1, http://string-db.org/) [12].

To begin with, the ensemble PPIN including 787896 interactions and 16730 genes were downloaded from the STRING database. Then, the intersection of interactions between the ensemble PPIN and microarray data were extracted. Finally, a new PPIN including 48778 interactions and 7953 genes was obtained.

### 2.4 DCN construction

For the patients undergoing CABG with sevoflurane treatment, there were two steps in the process of DCN construction.

Firstly, the construction of a binary co-expression network was implemented. To construct the binary gene co-expression network, edges were picked out according to the absolute value of the Pearson correlation coefficient (PCC) of the expression profiles of two genes. In order to take out the indirect correlation because of a third gene, the 1st order partial PCC was used [13]. Only edges whose correlations were greater than the pre-defined threshold δ were chosen. In our study, the δ value was set at 0.8, and thus maximal number of genes was connected in the DCN to be constructed.

Subsequently, weight values were assigned to the edges of the binary gene co-expression network on the basis of the P values of differential gene expression between the sevoflurane and control groups. In our analysis, we utilized one tail t-test to identify differential gene expression. The weight $w_{ij}$ on edge $(i,j)$ in the DCN was computed as:

$$w_{ij} = \begin{cases} 
(\log P_i + \log P_j)^{1/2}, & \text{if } \text{cor}(i,j) \geq \delta, \\
(2^{\max_{i,j} V} \log P_{ij})^{1/2}, & \text{if } \text{cor}(i,j) < \delta,
\end{cases}$$

In this formula, $P_i$ and $P_j$ were the respective P values of differential expression for gene i and gene j. $V$ was the...
node set in the DCN. cor(i,j) represented the absolute value of PCC between gene i and j according to their expression profiles. Under the weighting scheme, genes were assigned higher weight values when the genes were co-expressed and remarkably differentially expressed, which met our hypothesis that these genes were likely involved in a biological pathway that showed differential activities between the two groups.

2.5 Overview of the EgoNet algorithm for identification of human disease modules

In order to identify significant sub-networks that were functionally related with diseases, a novel method called EgoNet was proposed to exhaustively search and prioritize sub-networks and bio-markers from a large-scale biological network. The type of sub-network identified by our method was defined as module. Specifically, a module was a part of a network that involved a special node we were focusing on, which we called seed gene. Except seed gene, the network was made of a neighborhood including all nodes to which the seed gene was connected at a certain path length. The hypothesis of our model was that if the majority of neighbors of a central disease gene were disease genes, and the other neighbors were likely to be involved in the disease pathway. Alternatively, if most neighbors of the seed gene were related with a disease, the seed gene itself was regarded highly likely to exert a function in the disease. We planned to identify the hidden genes which exhibited no significance by themselves but were clustered in a sub-network in which genes collectively were highly predictive of disease state.

This algorithm took the PPIN and microarray data as input, and then EgoNet repeatedly scanned through all genes with two or more neighbors in the network. With each initial node (seed gene), it found the score of the level-one module based on how well the genes as a collection predicted the classification accuracy. Next, it spread outward from the seed gene progressively to involve more genes in the predictive model. The procedure stopped when the classification accuracy decreased. The process of growing module was known as snowball sampling. After getting the score of a module, the statistical significance was assessed by permutation test.

2.6 Identification of candidate modules

The objective of EgoNet is to extract significant modules from microarray profile and biological network data. The EgoNet algorithm consisted of three main steps: selection of seed genes, search of modules using snowball sampling, and refinement of modules.

2.6.1 Selection of seed genes

In the stage of seed genes selection, genes in the DCN were ranked using the topological measurement (degree) analysis. In brief, for DCN, Gk = (V, Ek) (1 ≤ k ≤ M) with an adjacency matrix Ak = (aijk)nxn, we constructed a function to denote the importance of vertex i:

\[ g(i) = \sum_{j \in N_k(i)} A_{ik} g(j) \]

In this function, Nk(i) stood for the set of neighbors of gene i in Gk; g(i) denoted the importance of vertex i in the DCN, Ak′ was the degree normalized weighted adjacency matrix which was computed based on \( A_k' = D^{-1/2} A_k D^{1/2} \) where D was diagonal matrix.

For each gene, after acquiring its ranks in DCN, marked as g = [g(1), . . ., g(M)], a z-score was computed for each rank g(l). Subsequently, the rank was obtained for that gene in the DCN via averaging the z-scores. The top 5% genes in the DCN were extracted as the seed genes.

2.6.2 Module search

This stage exhaustively searched all genes v whose addition caused the classification accuracy increase until there was no increase. For a given seed gene v ∈ V, it was treated as a module C={v}. For each vertex u in its neighborhood in the DCN, N(v) was defined as N(v) = UjNj(v) where Nj(v) was the neighbor set in Gi as the candidate for module C. For each u ∈ N(v), the classification accuracy increase was calculated between the new module C’ = CU{v} and C, that was to say, \( \Delta S(C', C) = S(C) - S(C') \). if \( \Delta S(C', C) \) was more than 0, which meant the addition of gene u improved the classification accuracy of the module C. The gene u whose addition maximizes \( \Delta S \) was added to the module C. If there were more than one node which were contained at each stage, we randomly selected one. The spreading stage stopped when the classification accuracy dropped. The above process of growing module was known as snowball sampling [14].

2.6.3 Refinement of candidate modules

In the refinement stage, the modules with classification accuracy smaller than 0.9 were removed. Moreover, the modules with gene size less than 5 were eliminated.
2.7 Statistical significance of the identified modules

A test of significance was implemented for the identified candidate modules to obtain the statistical significance. The null distribution of classification accuracy was obtained by randomly permuting 1000 times and computing the score from the same module each time. Then, the actual score of the module was indexed on the null distribution to get the P value which was calculated as the probability of the module having smaller classification accuracy by chance based on the following formula:

\[
P \text{ value} = \frac{\text{sum} \left( S(H_R) > S(H_C) \right)}{\text{count}(HR)}
\]

Where \( \text{count}(H_R) \) was the number of random permutation test, \( S(H_R) \) denoted the classification accuracy value of modules produced by random permutation test. \( S(H_C) \) stood for the classification accuracy value of candidate modules.

After obtaining P-values, the Benjamini & Hochberg [15] method was used to calculate the false discovery rate (FDR) to further adjust the P-values. In our study, FDR \( \leq 0.05 \) was considered as significant.

2.8 Pathway enrichment analysis

The development of diseases is frequently caused by altered pathways which participated in the biological process. For this, pathway enrichment analyses for differential module genes were performed. In the present analysis, all reference pathways in human were downloaded from Reactome database. Subsequently, the intersection of the genes in each reference pathway with PPIN genes was extracted. Then, a set of informative pathways were generated for subsequent analysis after removing pathways including less than 5 genes or more than 100 genes. Pathways with too few genes might not have sufficient biological information, and pathways with too many genes may be too generic. Next, genes in differential modules were aligned to each reference pathways to extract the pathways enriched by each differential module. Fisher’s exact test was used to evaluate the enrichment effects. Afterwards, we used Benjamini & Hochberg method to calculate the FDR to further adjust the P-values. In our study, pathways with FDR less than 0.05 were considered as the pathways enriched by a given module. Then, we ranked these pathways based on FDR values, and the pathway with the smallest FDR was selected as the significant pathway of a given module.

2.9 Power analysis to calculate the sample size

In the microarray profile of E-GEOD-4386, there were 4 groups (sevoflurane treatment group, baseline sevoflurane group, propofol treatment group, baseline propofol) which included 40 atrial samples collected at the beginning and at the end of the offpump CABG surgery. Among these 40 samples, 20 atrial samples were obtained from 10 patients receiving the anesthetic gas sevoflurane and 20 atrial samples were derived from 10 patients receiving the intravenous anesthetic propofol. In the current study, we utilized G*Power 3.1.9.2 software to make power analysis to calculate the sample size using \( \alpha = 0.05 \) and power(1-\( \beta \)) = 0.80 based on the results of NT-proBNP and the expression level of single seed genes.

Ethical approval: The conducted research is not related to either human or animals use.

3 Results

3.1 Construction of DCN and identification of candidate modules

To construct gene co-expression network, edges were selected according to the absolute value of the PCC of the expression profiles of two genes. A total of 2440 nodes were selected to construct the DCN based on the \( \delta \) greater than 0.8.

Based on the z-score distribution of 2440 nodes in DCN, we selected the genes with more than one connection and with the top 5% z-score value to serve as seed genes. Totally, 122 seed genes were identified. The top 20 seed genes are listed in Table 1. Subsequently, from every seed gene, we progressively grew the modules by levels. From every module, this procedure stopped when the classification accuracy dropped with the growth. After that, we extracted 122 modules and the mean node size in a module was 3. By setting the classification accuracy cutoff at 0.9 and the number of nodes in a module at 5, a total of 7 modules were identified, including module 80, 82, 83, 84, 85, 86 and 89.

3.2 Significance of candidate modules

After random permutation test with 1000 times, we obtained the FDR distribution of the identified modules, and we found that the FDR values of these 7 candidate...
Significant pathways in CABG induced by sevoflurane

We assumed that the ratio of sample size in the sevoflurane treatment group and control group was 1:1. Statistical power values suggested that the total samples were 36 based on the results of NT-proBNP. Moreover, statistical power values indicated that the total samples were respective 26, 30 and 26 based on the expression level of seed genes IL6, CCL20, and CXCL1. Significantly, in our study, the patients in sevoflurane treatment group and control groups were 10 and 10, respectively. According to these results, we thought the overall power value was relatively good, and the sample size was sufficient for our study.

### 3.3 Pathway enrichment analysis

After removing pathways with less than 5 genes or more than 100 genes, we ended up with 1136 pathways for enrichment analysis. After multiple test, we found that the genes of the module 80, 82, and 85 were all enriched in the pathway of chemokine receptors bind chemokines (FDR = 4.30E-06, 1.73E-06, 3.47E-04, respectively), the genes in the module 83 were enriched in the pathway of ERK2 activation (FDR = 5.04E-03). The specific information is shown in Table 3.

### 3.4 Power analysis

In the current study, we utilized G*Power 3.1.9.2 software to make power analysis to calculate the sample size using $\alpha=0.05$ and power(1-$\beta$) = 0.80 based on the results of NT-proBNP and the expression level of single seed gene (IL6, CCL20, and CXCL1).

We assumed that the ratio of sample size in the sevoflurane treatment group and control group was 1:1. Statistical power values suggested that the total samples were 36 based on the results of NT-proBNP. Moreover, statistical power values indicated that the total samples were respective 26, 30 and 26 based on the expression level of seed genes IL6, CCL20, and CXCL1. Significantly, in our study, the patients in sevoflurane treatment group and control groups were 10 and 10, respectively. According to these results, we thought the overall power value was relatively good, and the sample size was sufficient for our study.

### 4 Discussion

Former studies have demonstrated that sevoflurane, to a certain degree, plays cardioprotective role in CABG surgery [16,17]. However, the underlying mechanisms of
the effects of sevoflurane are still unclear. In the current study, in order to determine the protective molecular mechanisms of sevoflurane, the dataset E-GEOD-4386 was selected for further analysis. After snowball sampling, we extracted 122 candidate modules and the mean nodes in a module was 3. By setting the classification accuracy cutoff at 0.9 and the number of nodes in a module at 5, a total of 7 modules were identified, including module 80, 82, 83, 84, 85, 86 and 89. Significantly, these 7 modules were differential. Moreover, pathway analysis showed that genes in the module 80, 82, and 85 were all enriched in the pathway of chemokine receptors bind chemokines.

In the past decade, more effort has been done to extract differentially expressed genes (DEG) among different disease conditions, which can be employed as diagnostic signatures for classifying disease conditions or predicting clinical outcomes [18,19]. Nevertheless, bio-markers based on gene expression data alone are not reliable [20]. To solve this problem, network-based approaches are rapidly developed to obtain a comprehensive understanding of the disease process [21,22]. Network-based methods have many potential biological and clinical applications, including a better understanding of the effects of interconnection of disease genes and disease pathways, which, in turn, may offer better targets for drug development. So far, many methods have been created to integrate microarray profile with PPIN or pathway databases to identify key sub-network markers for predicting clinical outcomes [23,24]. For example, network-based Support Vector Machine (SVM) and Network-Guided Forests (NGF) have been utilized to select a set of genes which maximize the prediction performance [25,26]. However, the methods mentioned above are largely heuristic, and the definition of the sub-networks is ambiguous without a formal
Significant pathways in CABG induced by sevoflurane

449

Effects on cultured human kidney cells following ischemia reperfusion injury [40]. Several studies have also demonstrated that sevoflurane regulates the interaction of polymorphonuclear neutrophils and the microvascular endothelium, and this may play a key role in the initiation of reperfusion injury, even when cardiopulmonary bypass is performed [41,42]. Notably, in the current study, it was identified that the genes of the module 80, 82, and 85 were significantly enriched in the pathway of chemokine receptors bind chemokines. Thus, understanding the inflammation state of patients before CABG procedure is crucially important to decrease the post-operative complications.

Nevertheless, the sub-network markers involved in the molecular mechanisms underlying sevoflurane cardioprotective effects needs further confirmation at the gene and protein levels by means of reverse transcription quantitative polymerase chain reaction or western blot analyses. Hence, subsequent investigations are warranted, aiming to include a cell culture system or animal models to verify the results of our study. Taken together, sevoflurane might exert important functions in reducing myocardial reperfusion injury in patients following CABG, partially through regulating the pathway of chemokine receptors bind chemokines. However, the potential mechanisms for the effects of sevoflurane in cardioprotection need further confirmation.

Conflict of interests: Authors state no conflict of interest.

Abbreviations

coronary artery bypass graft: CABG
differential co-expression network: DCN
protein-protein interaction network: PPIN
Search Tool for the Retrieval of Interacting Genes: STRING
Pearson correlation coefficient: PCC
differentially expressed genes: DEG
Support Vector Machine: SVM
Network-Guided Forests: NGF

Reference


[25] Zhu Y., Shen X., Pan W., Network-based support vector machine for classification of microarray samples, BMC bioinformatics, 2009, 10, 1


[31] Lundeberg G.A., IN CELL RECRUITMENT, ACTIVATION, Chemokine Research Trends, 2007, 171


[38] Charo I.F., Taubman M.B., Chemokines in the pathogenesis of vascular disease, Circulation research, 2004, 95, 858-866


