**Full Length Article**

**GSEA analysis identifies potential drug targets and their interaction networks in coronary microcirculation disorders**

Nan Tang, Qiang Zhou, Shuang Liu, Huamei Sun, Haoran Li, Qingdui Zhang, Ji Hao, Chunmei Qi

*Department of Cardiology, The Second Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu 221000, China*

**A R T I C L E   I N F O**

**Keywords:**
- Coronary microcirculation dysfunction
- Gene set enrichment analysis
- Potential drug targets
- Interaction network
- Cardiovascular disease

**A B S T R A C T**

Coronary microcirculation dysfunction (CMD) is one of the main causes of cardiovascular disease. Traditional treatment methods lack specificity, making it difficult to fully consider the differences in patient conditions and achieve effective treatment and intervention. The complexity and diversity of CMD require more standardized diagnosis and treatment plans to clarify the best treatment strategy and long-term outcomes. The existing treatment measures mainly focus on symptom management, including medication treatment, lifestyle intervention, and psychological therapy. However, the efficacy of these methods is not consistent for all patients, and the long-term efficacy is not yet clear. GSEA is a bioinformatics method used to interpret gene expression data, particularly for identifying the enrichment of predefined gene sets in gene expression data. In order to achieve personalized treatment and improve the quality and effectiveness of interventions, this article combined GSEA (Gene Set Enrichment Analysis) technology to conduct in-depth research on potential drug targets and their interaction networks in coronary microcirculation dysfunctions. This article first utilized the Coremine medical database, GeneCards, and DrugBank public databases to collect gene data. Then, filtering methods were used to preprocess the data, and GSEA was used to analyze the preprocessed gene expression data to identify and calculate pathways and enrichment scores related to CMD. Finally, protein sequence features were extracted through the calculation of autocorrelation features. To verify the effectiveness of GSEA, this article conducted experimental analysis from four aspects: precision, receiver operating characteristic (ROC) curve, correlation, and potential drug targets, and compared them with Gene Regulatory Networks (GRN) and Random Forest (RF) methods. The results showed that compared to the GRN and RF methods, the average precision of GSEA improved by 0.11. The conclusion indicated that GSEA helped identify and explore potential drug targets and their interaction networks, providing new ideas for personalized quality of CMD.

**1. Introduction**

With the improvement of people’s material living standards and the change of living habits, the incidence rate of coronary microcirculation dysfunctions has gradually increased. As a cardiovascular disease, CMD involves the microcirculation of the heart and its pathological mechanism is complex [1-2]. Traditional treatment methods mostly focus on large vessel lesions, lacking drug intervention for microcirculation abnormalities. Exploring potential drug targets and their interaction networks for coronary microcirculation dysfunctions is key to developing new treatment strategies. With the mature development of bioinformatics, GSEA technology has made great progress and played an important role in clinical prediction, gene function analysis, and biological process regulation [3]. In the identification of potential drug targets and their interaction networks in coronary microcirculation dysfunctions, GSEA can provide effective gene expression data analysis, and on this basis, fully consider the global changes in the gene set to better interpret the complexity of microcirculation disorders. It has important practical value in promoting personalized development of coronary microcirculation dysfunction treatment and drug development.

With the rapid development and application of clinical medicine, CMD recognition has also achieved certain results. Crea Filippo believed that the onset of CMD involves a combination of functional and structural changes, which can lead to impaired coronary blood flow and myocardial ischemia. He provided the latest evidence on the potential...
pathophysiological mechanisms of CMD, with a particular focus on the role of cardiovascular risk factors and comorbidities, and discussed the specific pathogenic mechanisms of CMD in different cardiovascular diseases, laying the foundation for further research and development of new strategies for precision medicine methods [4]. McChord Johanna conducted a study on targeted drug therapy based on recognized disease mechanisms, and the results of the 2020 coronary heart disease microcirculatory dysfunction, reporting current knowledge on the mechanisms, and the results of the 2020 coronary heart disease microangiopathy regulated therapy trial showed improvements in angina symptoms and quality of life compared to the control group receiving standard treatment, highlighting the relevance of targeted drug therapy in this patient population [5]. Kunadian Vijay summarized the views of the expert group organized by the European Association for Percutaneous Cardiovascular Interventional Therapy and evaluated the importance of non-oncorrative coronary arteries. Based on existing research evidence and best clinical practices, he suggested that non-obstructive coronary artery spasm may be caused by various mechanisms, including coronary artery spasm and microvascular dysfunction, providing a definition and guidance for the diagnosis and management of CMD [6]. Del Buono Marco Giuseppe described the pathophysiological mechanisms of CMD and their mechanisms and prognostic roles in different cardiovascular diseases, and discussed recent clinical diagnostic models and potential treatment strategies, providing a basis for the study of CMD treatment methods [7]. Padro Teresa provided the latest evidence for the current understanding of the pathophysiological consequences of microvascular dysfunction, reporting current knowledge on the correlation between cardiovascular risk factors and comorbidity conditions of microcirculation dysfunction, as well as demonstrating its clinical relevance. He emphasized the clinical importance of CMD, which helped to stratify cardiovascular risk based on new concepts in precision medicine [8]. Although existing CMD identification and treatment methods have certain improvement capabilities for the prognosis of CMD, their research on potential drug targets and their interaction networks still cannot meet the current practical treatment needs.

The development of GSEA provides more possibilities for identifying potential drug targets and their interaction networks in CMD. Liu Jie conducted research on coronary artery disease recognition based on GSEA, defined four classifiers based on gene expression profiles in peripheral blood monocytes, and determined their potential in coronary artery disease detection. GSEA showed that differentially expressed genes are involved in the interaction network of coronary artery disease, and their expression results can effectively achieve non-invasive measurement [9]. To provide early diagnosis and treatment possibilities for coronary artery syndrome, Li Mingshuang revealed the key genes and transcription factors that affect the stability of coronary artery plaques through GSEA, and obtained the gene expression matrix of coronary artery syndrome patients and healthy subjects from public databases. The biological functions of differentially expressed genes were demonstrated through GSEA. Research has shown that three genes (vascular endothelial growth factor A, saposin secreted phosphoprotein 1, and vascular cell adhesion molecule 1) are involved in the molecular mechanism of coronary artery syndrome and can serve as biomarkers for disease progression [10]. Dong Feng believed that Takotsubo syndrome is a coronary microvascular disease, and used GSEA to perform deep pathway analysis and obtain the gene expression matrix for Takotsubo syndrome [11]. GSEA can fully consider the complexity of CMD pathophysiological mechanisms and the differences in gene expression, achieving scientific and objective analysis of potential drug targets and their interaction networks. However, most studies still have certain limitations in personalized treatment of CMD.

In order to improve the effectiveness and specificity of CMD treatment and improve patient intervention level, this article combined GSEA technology to study the identification of potential drug targets and their interaction networks in coronary microcirculatory dysfunctions. 60 disease samples, 60 healthy samples, and 967 gene data were collected from the Coremine medical database, GeneCards, and DrugBank databases, respectively. Ten gene signaling pathway datasets closely related to CMD were randomly selected as the target gene set. To verify the effectiveness of GSEA technology, this article conducted experimental analysis from four aspects: precision, ROC curve, correlation, and potential drug targets. In terms of precision, compared to GRN and RF methods, the average precision of GSEA has increased by 0.11; in terms of ROC curve, the specific AUC (Area Under Curve) value under the ROC curve of GSEA method reached 0.536, while the AUC values of GRN and RF method were 0.527 and 0.521, respectively; in terms of correlation and potential drug target analysis, the EDN1, ADRA1A, VWF, AGT, ATP2A2 gene sets had a higher degree of correlation with CMD. Through correlation analysis, it was found that the drugs corresponding to potential drug targets mainly included nicardipine, heparin, enalapril, tacrolimus, nitroglycerin, dobutamine, and acyclovir. The innovation of this article lies in the use of GSEA method to comprehensively and systematically analyze the gene expression data of CMD patients and healthy control groups. This global perspective helps to identify small but consistent changes in gene expression, as well as key biological processes and signaling pathways involved in CMD. And analyzed new drug targets to improve the efficiency of drug development, providing a theoretical basis for personalized treatment and precision medicine. In practical development, the GSEA method can effectively identify and explore potential drug targets in CMD, and conduct in-depth analysis of their interaction networks, which plays an important role in the sustained development of CMD treatment.

2. Identification of potential drug targets for coronary microcirculation dysfunctions and their interaction networks

CMD is an important factor in the occurrence and development of non obstructive coronary heart disease. Cardiovascular events caused by CMD seriously endanger the health of patients and have a great impact on their daily lives. Targeted treatment for CMD is currently an urgent clinical problem to be solved [12]. The onset of CMD is a complex process involving multiple factors and multi-level interactions [13-14]. In addition to the problems inherent in blood vessels, they are also related to factors such as vasoconstriction function, metabolism, thrombosis, inflammation, neural regulation, and endothelial dysfunction. The mutual influence between various elements makes their pathological mechanisms more complex and diverse.

GSEA is a technology that utilizes genomic information for bioinformatics analysis [15]. It can reveal the degree of gene enrichment in a specific biological environment within the organism, interpret gene expression data, and identify pathways and functional modules related to life activities [16-17]. This article aims to analyze the distribution of genes on different gene sequences through GSEA and obtain the enrichment scores of gene sets. By using hypothesis testing models for statistical analysis, effective identification and exploration of potential drug targets and their interaction networks in CMD can be achieved, providing new ideas and methods for the study of the pathogenesis and clinical treatment of CMD.

2.1. Data collection

(1) Database

The molecular biology and genetics research of CMD has established corresponding disease-gene association data. This article combines public databases to collect CMD related gene data. The public databases used in this article mainly include: Coremine medical database, GeneCards, and DrugBank. Among them, the Coremine medical database belongs to the medical database, which includes relevant information from multiple disciplines such as biomedicine, clinical practice, pharmacology, biochemistry, etc. It mainly provides comprehensive medical
literature information and medical databases. GeneCards, as an open gene database, integrates gene data resources from 125 databases to provide adequate and comprehensive data information for research on human genome, transcriptomics, proteomics, clinical genetics, and other functions. DrugBank is the world’s largest database of drugs and drug targets. This database integrates detailed information from pharmaceutical chemistry, pharmacology, and the pharmaceutical industry, providing a basis for research on the sequence, structure, and metabolic pathways of drug targets.

(2) Collection of relevant gene data

The collection of CMD related gene data is shown in Fig. 1.

Given the complex pathogenesis, pathological classification, and diverse nature of CMD, this article first uses Coronary Microcirculation Dysfunction as the unified keyword. According to the genetic data collection steps in Fig. 1, topic words closely related to CMD are retrieved and annotated through a medical topic dictionary, and finally CMD related topic words and corresponding identification numbers are determined.

Based on the retrieved keywords, the genes associated with them are searched in the GeneCards database, and the disease gene association and gene names are recorded. Based on the correlation between diseases and genes and the number of literature in the database, the “disease-gene” association table is obtained.

In the Coremine Medical database and DrugBank database, relevant literature on this disease and gene is loaded, and keywords are searched in the title and abstract. Literature related to genes are placed in the corresponding files. If no keywords are found in the entire text, these materials are placed in unrelated files. The selected literature is classified and sorted in the order of their publication period. On this basis, phenotype-genotype relationships related to CMD are identified based on the identified CMD related keywords, and the data results retrieved from the DrugBank database are correlated. The obtained data samples are divided into two categories: disease samples and health samples. The training set consists of human drug target data, non drug target data, CMD drug target data, and CMD non drug target data.

When establishing a disease sample set, the first step is to obtain drug target data that has been confirmed by research, and the target sequence data is mainly downloaded from the DrugBank database. The obtained target sequence data is divided into two fields: one is the target number on DrugBank, and the other is the target name on the gene label. On this basis, non human protein sequences are screened out through computer programs and used as a set of disease samples.

In the construction of a healthy sample set, first, all drug target sequences are downloaded from the DrugBank database through a computer, and non human protein sequences are removed. Its screening is mainly based on the homology of protein sequences. If the protein sequences in the sample do not have or only have a very low degree of similarity with the protein sequences in the human protein database, usually exhibiting low homology or low conservation, then these protein sequences are considered unrelated to humans. This step is roughly the same as the one performed to construct a disease sample set. On this basis, all human protein sequences are obtained from the GeneCards database, excluding previously obtained protein sequences and their family data. These proteins are subjected to novel genome sequencing, and data is selected from non drug targets screened and assembled into a healthy sample set.

2.2. Data processing

To avoid errors caused by differences in conditions and different steps in protein sequence data extraction and processing, filtering methods are used to standardize the data, screen and remove genes that do not have an effect on the target phenotype or state, and minimize the statistical testing of these gene data. These genes display low signal strength and may be considered unrelated to the disease of interest.

Fig. 1. Gene data collection

Fig. 2. Gene expression data filtering and processing
intensity or amplitude in all samples, making them less likely to participate in the target phenotype or state. Its processing is shown in Fig. 2.

GSEA is different from single gene analysis in that it considers the enrichment of predefined gene sets in the entire gene expression data. Through this method, it is possible to comprehensively evaluate the overall changes in gene sets, rather than the significance of individual genes. This global perspective can reveal complex biological processes related to diseases. The integrated analysis combined with other omics data such as proteomics provides a more comprehensive biological landscape, which helps to gain a deeper understanding of the molecular mechanisms of diseases. According to the gene expression data filtering method in Fig. 2, with a gene count of N and a sample count of M, there are two phenotypes in the sample set: positive phenotype and negative phenotype. These two phenotypes are set to be $G_1$ and $G_2$, respectively, with a sample count of n1 and n2 for both phenotypes. Given a signaling pathway and representing it as $L_0$, the number of genes included in this pathway is set to $R$ [18-19]. Firstly, based on the interaction information between genes, the correlation degree of each gene is statistically analyzed to obtain the frequency $f_{o}$ of genes. Then, the frequency of each gene appearing in all pathways is statistically analyzed, and the frequency $f_{r}$ of the reverse pathway is calculated. The frequency $f_{o}$ of the gene is integrated with the frequency $f_{r}$ of the reverse pathway, and the degree of gene enrichment in the interaction network is finally calculated.

Pathways involve a large number of gene and molecular interactions, and the construction of gene sets is usually based on certain assumptions or definitions. Therefore, some genes related to pathways may not be included in the gene set due to their unclear mode of action or importance in a specific context. Due to the possibility of certain genes not being included in the gene set in the pathway, the gene set of gene expression data is subjected to intersection processing, and the correlation between the genes at their intersection and phenotype is calculated. This step is expressed as using the formula [20]:

$$C_r = \frac{\text{mean}(G_{11}) - \text{mean}(G_{22})}{\text{std}(G_{11}) - \text{std}(G_{22})}$$ (1)

Calculating the correlation between genes and phenotype can help identify potential genes related to phenotype. By analyzing genomic data, it is possible to determine which genes are associated with the occurrence or expression of specific phenotypes, thereby identifying candidate genes. Among them, mean represents the mean expression value of each phenotype gene, and std represents its standard deviation. A list containing $\alpha$ genes is generated in descending order of association, and this list is represented as M. When encountering a gene in pathway $L_0$, the enrichment score is correspondingly increased; on the contrary, when encountering genes that are not in pathway $L_0$, the enrichment score is reduced, and the specific formula is [21-22]:

$$E_{o}(L_0) = \max_{1 \leq k \leq M} \left| \sum_{i \neq j|k \neq i} \left( \frac{|C_{ij}(1 + I_{ij})|}{\sum_{i \neq j|k \neq i} C_{ij}(1 + I_{ij})} \right) - \sum_{i \neq j|k \neq i} \frac{1}{|\alpha - R|} \right|$$ (2)

Based on the score changes during this process, the enrichment score curve is obtained. The maximum distance from a point on the curve to the abscissa is $E_{o}(L_0)$.

Genes are randomly substituted $N$ times, that is, $R$ genes are randomly selected from the list M as pathway genes. Then, the pathway enrichment degree after permutation is calculated to be $E_{o}(L_0)$, and finally the number of $|E_{o}(L_0)| > |E_{r}(L_0)|$ is counted as $N_R$. The significance $p$-value is the ratio of $N_R$ to $N$.

Computers cannot analyze and arrange amino acid residues in drug target data, so their order needs to be converted into recognizable numerical sequences. This requires extracting features from protein sequences based on their amino acid sequences and chemical and physical properties as feature spaces [23], as shown in Table 1.

In Table 1, the composition, transfer, and distribution coefficients belong to autocorrelation coefficients, which are calculated independently using statistical methods.

Assuming the content of amino acids in the sequence is $f(A)$, the calculation is expressed as [24-25]:

$$f(A) = \left( \frac{T_A}{l} \right) \times 100\%$$ (3)

Assuming the content of dipeptides in the sequence is $f(D)$, its calculation is expressed as:

$$f(D) = \left( \frac{T_A}{l(l-1)} \right) \times 100\%$$ (4)

The variable meanings of Formulas (3) and (4) are shown in Table 2. Autocorrelation features refer to the degree of correlation between the special structure and chemical and physical properties of proteins to a certain extent [26]. On this basis, feature extraction is achieved through autocorrelation feature calculation. Firstly, it is necessary to normalize the chemical and physical properties of the obtained amino acids. The specific formulas are expressed as:

$$w'_A = \frac{(w_A - \bar{w})}{\sigma}$$ (5)

$$\bar{w} = \frac{\sum_{A=A}^{D} w_A}{D}$$ (6)

$$\sigma = \sqrt{\frac{1}{D-1} \sum_{A=A}^{D} (w_A - \bar{w})^2}$$ (7)

Among them, $w_A$ represents the chemical and physical properties of amino acids before normalization, and $w'_A$ represents the chemical and physical properties of amino acids after normalization.

In attribute components, the transfer and distribution autocorrelation characteristics refer to the distribution patterns of specific structures or physicochemical properties of amino acids on protein sequences. The formula for the composition coefficient is expressed as [27]:

$$C_\alpha = \frac{\alpha}{T}$$ (8)

The formula for the transfer coefficient is defined as:

$$C_\alpha = \frac{\alpha_o}{T - 1}$$ (9)

The distribution coefficient describes each attribute in the sequence and is defined by the formula [28-29]:

| Table 1 |
|-----------------|-----------------|-----------------|
| **Protein feature space** | **Sequence** | **Feature** |
| 1 | Amino acid | A |
| 2 | Dipeptide | D |
| 3 | Composition coefficient | $C_\alpha$ |
| 4 | Transfer coefficient | $C_\alpha$ |
| 5 | Distribution coefficient | $C_\alpha$ |

| Table 2 |
|-----------------|-----------------|-----------------|
| **Sequence** | **Variable** | **Meaning** |
| 1 | $T_A$ | The total number of amino acids |
| 2 | $l$ | Sequence length |
| 3 | $r$ | $r$-type dipeptide |
| 4 | $s$ | $s$-type dipeptide |
| 5 | $T_{rs}$ | The total number of $r$-type dipeptides and $s$-type dipeptides |
Table 3
Gene expression data samples and number of genes

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Item</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of disease samples</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>Number of healthy samples</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>Number of genes</td>
<td>967</td>
</tr>
</tbody>
</table>

\[
C_i = \frac{1}{1-(\theta_0 + \theta_s)^2} \tag{10}
\]

On this basis, the data is classified, and the gene expression data of the pathway gene set is combined with the phenotype information of the sample. The corresponding classification model is established, and the area below the ROC curve and the coordinate axis is calculated as the AUC (Area Under Curve) value, which ranges from 0.5 to 1. The larger the AUC value, the better the understanding of the interaction between the expression values of the genes in this pathway, and the higher the correlation with the onset of CMD.

Based on the correlation results, genes closely related to the development of CMD are screened, and potential drug targets are identified. In determining potential drug targets, three conditions need to be met: firstly, the drug corresponding to the target is a non experimental drug; secondly, the pathway in which the target protein is located is highly associated with CMD disease; thirdly, the drug information corresponding to the target needs to be complete.

3. GSEA experiment results

To verify the effectiveness of GSEA in identifying potential drug targets and their interaction networks in coronary microcirculation dysfunctions, this study conducted experimental analyses from four aspects: precision, ROC curve, correlation, and potential drug targets.

3.1. Data collection results

During the data collection process, this article collected gene expression data using the Coremine medical database, GeneCards, and DrugBank as databases. (1) Coremine medical database

2015 disease-gene relationships were obtained from the Coremine medical database, with each disease-gene relationship representing a gene data. Among them, 983 were reviewed and confirmed to be unrelated to CMD, and 1032 were confirmed to be related to CMD. After deduplication, 74 were removed, and ultimately 958 gene data were obtained.

(2) GeneCards database

102 disease phenotype-genotype records were obtained from the GeneCards database, containing a total of 76 gene data. Among them, 42 genes were duplicated with 958 related genes obtained from Coremine medical database; 25 were duplicated with unrelated genes obtained from Coremine medical database, and the remaining 9 were not included in the 2015 records of Coremine medical database. Therefore, they were included.

(3) DrugBank database

In the DrugBank database, 775 drug target sequences related to CMD were obtained, and disease sample sets and healthy sample sets were constructed based on drug target data.

The relevant genes obtained from the Coremine medical database and GeneCards were integrated, and the sample set from the DrugBank database was screened. The final gene expression data samples and the number of genes obtained are shown in Table 3.

From Table 3, it can be seen that in the public database, the final number of disease samples obtained was 60; the number of healthy samples was 60; the number of genes was 967.

Table 4
Target gene set

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Gene set</th>
<th>Related gene</th>
<th>Describe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EDN1</td>
<td>Endothelin 1</td>
<td>Encoding angiotensin, involving vasodilation and contraction</td>
</tr>
<tr>
<td>2</td>
<td>NOS3</td>
<td>Nitric oxide synthase 3</td>
<td>Related to the production of nitric oxide, which may affect vascular function</td>
</tr>
<tr>
<td>3</td>
<td>KCNMB1</td>
<td>Potassium Calcium-Activated Channel Subfamily M Regulatory Beta Subunit 1</td>
<td>Related to ion channels in vascular smooth muscle cells</td>
</tr>
<tr>
<td>4</td>
<td>ADRA1A</td>
<td>Adrenaline a 1a receptor</td>
<td>Participate in angiotensin regulation</td>
</tr>
<tr>
<td>5</td>
<td>VWF</td>
<td>Von Willebrand factor</td>
<td>Regulated the generation of angiotensin</td>
</tr>
<tr>
<td>6</td>
<td>ACE</td>
<td>Angiotensin converting enzyme</td>
<td>Regulating the generation of angiotensin</td>
</tr>
<tr>
<td>7</td>
<td>AGT</td>
<td>Angiotensinogen</td>
<td>Precursor molecules of angiotensin</td>
</tr>
<tr>
<td>8</td>
<td>GUCY1A3-GUCY1B3</td>
<td>Guanylate cyclase</td>
<td>Related to nitric oxide signaling pathway</td>
</tr>
<tr>
<td>9</td>
<td>THBD</td>
<td>Thrombomodulin</td>
<td>Plays a regulatory role in vascular endothelial cells</td>
</tr>
<tr>
<td>10</td>
<td>ATP2A2</td>
<td>Calcium ion transport ATPase A2</td>
<td>Participate in the calcium ion cycle of myocardial cells</td>
</tr>
</tbody>
</table>

Table 5
GSEA results

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Gene set</th>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EDN1</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>NOS3</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>KCNMB1</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>ADRA1A</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>VWF</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>ACE</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>AGT</td>
<td>24</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>GUCY1A3-GUCY1B3</td>
<td>30</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>THBD</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>ATP2A2</td>
<td>19</td>
<td>11</td>
</tr>
</tbody>
</table>

3.2. Data analysis results

The gene data of disease samples and health samples were tested, and 102 gene signaling pathways were extracted from the GSEA website for different gene data and screened. Among these 102 gene signaling pathways, a total of 54 were datasets that can be determined to be closely related to CMD, which means that these 54 gene sets should have significant differences in expression between healthy samples and CMD disease samples. In these datasets closely related to CMD, this article extracted 10 datasets as the target gene set for experimental analysis, as shown in Table 4.

(1) Precision

The gene set selected in Table 4 was used as the object, and precision was used as the indicator. The effectiveness of GSEA was evaluated and compared with the commonly used GRN and RF methods. The precision calculation formula is expressed as:

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{11}
\]

Among them, TP represents the number of gene sets that can be screened by the method and are in the target gene set, while FP represents the number of gene sets that are screened by the method but do not exist in the target gene set. Based on the measurement criterion of
The precision, the TP and FP results of the three methods are shown in Tables 5 to 7. From the precision results in Fig. 3, it can be seen that GSEA had a significant advantage in precision under different gene sets. In GSEA, its highest precision value reached 0.68, and its average precision in screening different gene sets was about 0.61; in GRN analysis, its highest precision and mean in gene set analysis were 0.55 and 0.50, respectively; the highest and average precision values of RF in gene set analysis were 0.54 and 0.50, respectively. From the specific comparison results, compared to GRN and RF methods, the average precision of GSEA improved by 0.11. Both GRN and RF algorithms study gene expression differences between different phenotypes from both point and edge perspectives, while GSEA can balance gene expression intensity and interaction, thereby more precisely screening differential gene sets.

(2) ROC curve

In order to further comprehensively analyze the recognition performance of GSEA in potential drug targets and their interaction networks in CMD, this paper calculated the specific and sensitive ROC curves and AUC values under the ROC curves based on the gene expression data and sample phenotype data corresponding to each pathway gene set. The analysis results of different methods on CMD data are shown in Fig. 4.

From Fig. 4, it can be seen that in gene data analysis, the AUC value under the ROC curve of GSEA method was larger, with a specific result of 0.536. The AUC values of GRN method and RF method were 0.527 and 0.521, respectively. From this result, it can be seen that compared to the

### Table 6

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Gene set</th>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EDN1</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>NOS3</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>KCNMB1</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>ADRA1A</td>
<td>34</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>VWF</td>
<td>40</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>ACE</td>
<td>46</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>AGT</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>8</td>
<td>GUCY1A3-GUCY1B3</td>
<td>31</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>THBD</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>ATP2A2</td>
<td>38</td>
<td>34</td>
</tr>
</tbody>
</table>

### Table 7

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Gene set</th>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EDN1</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>NOS3</td>
<td>37</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>KCNMB1</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>ADRA1A</td>
<td>41</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>VWF</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>ACE</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>AGT</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>GUCY1A3-GUCY1B3</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>THBD</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>ATP2A2</td>
<td>20</td>
<td>18</td>
</tr>
</tbody>
</table>

According to the TP and FP results in Tables 5 to 7, the precision of the three methods was calculated. The results are shown in Fig. 3.
other two methods, the GSEA method had a more ideal gene set analysis ability, which can more precisely reveal the interactions between drug targets, thereby improving the reliability of gene set screening and target recognition.

(3) Correlation

Correlation analysis can help identify potential drug targets, namely genes associated with CMD. By analyzing the correlation between genes and CMD phenotype, it is possible to determine which genes may be key factors affecting the occurrence and development of CMD, thereby providing targets for the development of new treatment methods. In order to screen out the gene set most affected by drugs, this study analyzed the differential enrichment of each gene set under various treatment conditions through GSEA, and calculated the enrichment score for normalization, effectively identifying the gene set with the highest correlation with potential drug targets in CMD. The normalized enrichment score (NES) results are shown in Fig. 5.

The normalized enrichment score ranged from -1 to 1. Generally speaking, positive values indicate significant clustering of gene sets under a given conditional pathway. From Fig. 5, it can be seen that the NES results of the target gene set selected in this article were all positive, and there were certain differences in NES results under different gene sets. Among them, the NES results of EDN1, ADRA1A, VWF, AGT, and ATP2A2 reached above 0.5, with specific score results of 0.62, 0.56, 0.71, 0.58, and 0.73, respectively. This indicates that these gene sets have a higher degree of correlation with CMD and have a more significant impact in the interaction network, playing an important regulatory role in CMD.

To objectively analyze the NES results, this article conducted statistical analysis using a hypothesis testing model. The results are shown in Table 8.

From the statistical analysis results in Table 8, it can be seen that the p-value results of EDN1, ADRA1A, VWF, AGT, and ATP2A2 with NES results above 0.5 were relatively low in the statistical analysis results, indicating that the analysis results of GES were relatively objective.

(4) Potential drug targets

Based on the enrichment score results and statistical analysis results
in the correlation analysis, it was believed that EDN1, ADRA1A, VWF, AGT, ATP2A2 had a high correlation with CMD. This article analyzed the protein sequences and potential drug targets associated with them in the gene set. The results are shown in Table 9.

Based on the 16 potential drug targets in Table 9, by associating them with the downloaded drug target sequences in the DrugBank database, 6 experimental drugs and 3 drugs with incomplete information were excluded, resulting in 7 potential drug targets. The corresponding basic information of existing drugs is shown in Table 10.

In Table 10, the drugs corresponding to potential drug targets mainly include nicardipine, heparin, enalapril, tacrolimus, nitroglycerin, dobutamine, and acyclovir, which may have certain potential effects in the treatment of CMD.

4. Discussion

To analyze the effectiveness of studying potential drug targets and their interaction networks for coronary microcirculation dysfunctions under GSEA, this study conducted experimental analysis from four aspects: precision, ROC curve, correlation, and potential drug targets.

(1) Precision experimental analysis results

From the precision analysis results, compared to the GRN method and RF method, the average precision of GSEA has improved by 0.11. Both GRN and RF methods focus on analyzing the expression differences of a single gene or gene pair, while ignoring the overall attributes and correlations of the gene set. GSEA enables comprehensive reflection of the overall characteristics of genes, and can arrange genes in order according to their pathways, thereby reflecting the degree of gene enrichment under different phenotypes.

(2) ROC curve experimental analysis results

In the ROC curve experimental analysis, the AUC value under the ROC curve of the GSEA method reached 0.536. Compared to GRN and RF methods, GSEA method has more significant advantages in analyzing reliability. Through effective analysis of gene sets, GSEA can gain a more adequate and comprehensive understanding of the CMD interaction network, and effectively identify gene sets with greater impact.

(3) Analysis results of correlation experiment

Under GSEA, it was found that the NES results of EDN1, ADRA1A, VWF, AGT, and ATP2A2 were higher, indicating a higher influence in the interaction network. This suggests that these gene sets may be key regulatory factors in the pathogenesis of CMD, with significant potential impact on the development of CMD and personalized treatment.

(4) Analysis results of potential drug target experiments

Based on the analysis of potential drug target experiments, experimental and incomplete drugs were excluded, and ultimately 7 potential drug targets were retained. The corresponding drugs mainly include nicardipine, heparin, enalapril, tacrolimus, nitroglycerin, dobutamine, and acyclovir. In the subsequent development of clinical treatment for CMD, in-depth research on these drugs may help improve the quality and level of CMD treatment.

5. Conclusions

With the increasing incidence rate of cardiovascular diseases, the treatment of CMD has attracted more and more attention. The pathogenesis of CMD is complex, and traditional treatment methods lack specificity, making it difficult to achieve differentiated treatment of CMD. In order to improve the quality of CMD treatment and improve patient prognosis, this article combined GSEA to conduct in-depth research on potential drug targets and interaction networks in CMD. The GSEA method has not only effectively improved the precision of gene data analysis, but also significantly improved the reliability of CMD interaction network analysis, achieving effective identification and mining of potential drug targets. The final results indicated that nicardipine, heparin, enalapril, tacrolimus, nitroglycerin, dobutamine, and acyclovir could be potential drugs for the quality of CMD. Further analysis of their pharmacological effects may help improve the therapeutic effect of CMD. This article analyzed and identified potential drug targets and their interaction networks in CMD through GSEA. Although it provides some guidance for promoting personalized treatment of CMD, there are still limitations in this article. GSEA analysis relies on high-quality gene expression data. The sample size in the experimental analysis section of this article is not sufficiently rich, and GSEA mainly focuses on the enrichment of gene expression. This article did not consider factors such as epigenetic regulation and non-coding RNA. In future research, improvements would be considered from the perspective of biological information quality and sample sources to continuously promote the high-quality development of CMD treatment.

Funding

This work was supported by the Xuzhou Municipal Health Commission Youth Medical Research Project (Grant No. XWKYHT20210566) and Xuzhou Medical University Affiliated Hospital of Science and Technology Development Excellent Talent Fund Project (Grant No. XYPY202249).

CRediT authorship contribution statement

Nan Tang: Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. Qiang Zhou: Data curation, Formal analysis, Investigation. Shuang Liu: Project administration, Resources. Huamei Sun: Methodology, Resources. Haoran Li: Resources, Validation. Qingdui Zhang: Validation, Writing – review & editing. Ji Hao: Validation, Visualization. Chunmei Qi: Writing – original draft, Writing – review & editing.

Declaration of competing interest

There are no potential competing interests in my paper.

References


[10] Li Mingshuang, Ren Conglin, Wu Chemia, Li Xinyao, Li Xinyi, Mao Wei. Bioinformatics analysis reveals diagnostic markers and vital pathways involved in
N. Tang et al.

slas technology xxx (xxxx) xxx

9


[26] Shuang Liu was born in Ruzhou, Henan, P.R. China, in 1998. He received the bachelor's degree from Hubei Medical College, P.R. China. Now, he is studying in the Graduate School of Xuzhou Medical University. His research interests include coronary heart disease, angina and coronary microcirculation disorders. Email: lsz@stju.edu.cn

[27] Shaoming Lian was born in Xuzhou, Jiangsu, P.R. China, in 1973. He received the Master degree from Xuzhou Medical University, P.R. China. Now, he works in the Second Affiliated Hospital of Xuzhou Medical University. His research interests include coronary heart disease, heart failure and myocardial infarction. Email: sjj1973@163.com

[28] Qingdai Zhang was born in Xuzhou, Jiangsu, P.R. China, in 1979. He received the doctor’s degree from Nantong University, P.R. China. Now, he works in the Second Affiliated Hospital of Xuzhou Medical University. His research interests include coronary heart disease, angina and coronary microcirculation disorders. Email: zdq@nust.edu.cn

[29] Ji Hao was born in Xuzhou, Jiangsu, P.R. China, in 1979. He received the doctor’s degree from Dongnan University, P.R. China. Now, he works in the Second Affiliated Hospital of Xuzhou Medical University. His research interests include coronary heart disease, cardiac intervention and microcirculation. Email: hqj2022@163.com

Nan Tang was born in Jiuxiang, Jiangxi, P.R. China, in 1996. He received the bachelor’s degree from Jiujiang University, P.R. China. Now, he is studying in the Graduate School of Xuzhou Medical University. His research interests include coronary heart disease, angina and coronary microcirculation disorders. E-mail: sjq202122@163.com

Qi Yang was born in Xuzhou, Jiangsu, P.R. China, in 1998. He received the Master degree from Xuzhou Medical University, P.R. China. Now, he works in the Second Affiliated Hospital of Xuzhou Medical University. His research interests include coronary heart disease, angina and coronary microcirculation disorders. Email: lsz@stju.edu.cn

Huaimei Sun was born in Xuzhou, Jiangsu, P.R. China, in 1984. She received the Master degree from Zunyi Medical College, P.R. China. Now, she works in the Second Affiliated Hospital of Xuzhou Medical University. Her research interests include coronary heart disease, heart failure and myocardial infarction. Email: shm1919@163.com

Haoran Li was born in Xuzhou, Jiangsu, P.R. China, in 1996. He received the Master degree from Xuzhou Medical University, P.R. China. Now, he works in the Second Affiliated Hospital of Xuzhou Medical University. His research interests include coronary heart disease, heart failure and myocardial infarction. Email: 15005206607@163.com

Qi Yang was born in Xuzhou, Jiangsu, P.R. China, in 1996. He received the Master degree from Xuzhou Medical University, P.R. China. Now, he works in the Second Affiliated Hospital of Xuzhou Medical University. His research interests include coronary heart disease, angina and coronary microcirculation disorders. Email: lsz@stju.edu.cn

Chunmei Qi was born in Shenyang, Liaoning, P.R. China, in 1969. She received the doctor’s degree from Dongnan University, P.R. China. Now, she works in the Second Affiliated Hospital of Xuzhou Medical University. Her research interests include coronary heart disease, cardiac intervention and microcirculation. Email: 760020220858@xzhmu.edu.cn