Screening differentially expressed genes and the pathogenesis in atopic dermatitis using bioinformatics

Pengfei Yang, Yanmei Li, Tingbao Li, Liang Yuan, Sinong Wang*
Clinical College of Chinese Medicine, Gansu University of Chinese Medicine, Lanzhou, Gansu Province, 730000, China

ARTICLE INFO
Original paper
Article history:
Received: July 14, 2023
Accepted: November 26, 2023
Published: December 31, 2023

Keywords:
AD, bioinformatics, differential expression, genes, pathogenesis, inflammatory response

ABSTRACT
Atopic dermatitis (AD) is a common chronic skin inflammation. It was to screen differentially expressed genes (DEGs) and related biological functional pathways in atopic dermatitis (AD) by bioinformatics methods, and to understand the pathogenesis of AD. Nine gene chip datasets GSE120721 and GSE32924 in the public database NCBI Gene Expression Omnibus (GEO) were adopted. Differential expression analysis between the patient group and controls was performed by applying the zero-code differential expression analysis tool GEO2R, and a few DEGs were screened. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out. In addition, STRING online database was employed to predict the potential relationship among DEGs, the protein-protein interaction network (PPI) was drawn, and the module analysis of PPI was performed using the Cytoscape plugin MCODE. 233 DEGs were screened out, including 134 up-regulated genes and 99 down-regulated genes. GO analysis suggested that these DEGs were mainly involved in biological processes (BP), cellular components (CC), and molecular function (MF). KEGG analysis displayed that these DEGs were mainly involved in NF-kappa B signaling, cell cycle, T cell receptor signaling, and other pathways. PPI analysis indicated that there were complex interactions among DEGs, and module analysis further revealed the important roles of DEGs in regulating immune response, inflammatory response, and skin barrier function. The above findings provide a valuable reference for the development of new treatment options.

Introduction

Definition, epidemiology, and clinical manifestations of AD
AD characterized by itchy, red, dry, and broken skin is a disease that is widely distributed throughout the world, especially in developed countries and urban areas (1-3). The global prevalence of AD is estimated at 10-20% in children and 1-3% in adults and is increasing in many regions. As a result, AD has become a public health problem. The etiology and pathogenesis of AD are very complex, including genetic, immune, and environmental factors. Many studies suggested that the pathogenesis of AD is correlated with the abnormal activation of varieties of immune cells, cytokines, and signaling pathways (4-6). In addition, environmental factors such as allergens, microbial infection, irritant substances, and climate change may also contribute to the pathogenesis of AD. Traditional treatments for AD include topical agents on the skin and oral antihistamines, but these have limited effectiveness and may lead to side effects and drug resistance (7,8). Therefore, understanding AD pathogenesis and exploring new therapeutic targets have great significance for the treatment and prevention of AD.

Application of bioinformatics in DEGs screening
The development and application of bioinformatics technology have provided a new way for the study of AD. Bioinformatics is a high-throughput technology based on big data, which can help researchers mine related DEGs from many gene expression data (9-11). Specifically, the adoption of bioinformatics in DEGs screening included the following steps.

(i) Data cleaning and preprocessing: The acquired transcriptome data were pretreated, including data denoising, normalization, standardization, and batch effect elimination, etc., to ensure the accuracy and reliability of the data.

(ii) Differential expression analysis: the transcriptome data of the screened disease and the normal groups were analyzed, and the DEGs related to the disease were screened by comparing the two groups of data.

(iii) Functional annotation: bioinformatics analysis and annotation were carried out on the selected DEGs to determine their biological functions and metabolic and signaling pathways.

(iv) Pathway analysis: DEGs were input into bioinformatics software, and pathway analysis was performed to explore the functions and roles of these genes in metabolic and signaling pathways, and BP.

The application of bioinformatics in DEGs screening helps to identify DEGs and metabolic pathways related to the pathogenesis of AD, reveals the pathogenesis and pathophysiological process of AD, and provides new targets and strategies for the treatment and prevention of AD (12). For example, in the study of AD, bioinformatics has found multiple genes and metabolic pathways related to the pathogenesis of AD, such as immunomodulatory genes,

* Corresponding author. Email: taipo5385747@163.com
Copyright: © 2023 by the C.M.B. Association. All rights reserved.
http://dx.doi.org/10.14715/cmb/2023.69.15.12

Cellular and Molecular Biology, 2023, 69(15): 73-78
inflammatory pathways, and keratinization pathways (13-15). These results provide new directions and strategies for the treatment and prevention of AD and help to improve the quality of life and health of AD patients.

**Purpose and significance**

The main objective was to screen out DEGs related to the pathogenesis of AD and explore their possible pathogenesis using bioinformatics technology. Specifically, through large-scale transcriptome data in public databases, gene expression data in different disease states were differentially analyzed by bioinformatics analysis methods, and functional annotation and pathway analysis were performed. Finally, a batch of DEGs and related BP and pathways related to the pathogenesis of AD were obtained.

Firstly, AD transcriptome data in public databases were screened and pre-processed, including data cleaning, normalization, quality control, and other steps. Next, bioinformatics software was adopted for DEGs screening and functional annotation to identify DEGs related to the pathogenesis of AD and to explore the BP and pathways in which these genes may be involved. Finally, the selected DEGs were verified by experimental verification. Through these analyses, we can further understand the pathological mechanism of AD and provide new directions and strategies.

In conclusion, bioinformatics technology was adopted to analyze large-scale transcriptome data and explore the possible pathogenesis. It aimed to present new perspectives and ideas for AD research.

**Materials and Methods**

**Data sources**

Data source mainly comes from the public database NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/). These databases contain a large amount of bioinformatics data, including gene expression data, genomic data, and clinical data. According to the inclusion criteria in Table 1, two eligible gene microarray data sets were found, and the AD-related gene chip data set GSE120721 was downloaded. The data set contained 52 samples, having 22 normal samples and 30 AD patients. The microarray platform was GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array was adopted for GSE32924, which consisted of 33 samples, with 8 normal subjects and 25 AD patients.

**DEGs screening**

Differential expression tools GEO2R (https://www.ncbi.nlm.nih.gov/geo/) with zero code was employed for GSE28146 DEGs of microarray data analysis. GEO2R is an application in the GEO2 online database that identifies DEGs in 2 or more samples and ranks genes importance to compare the gene in AD patients and controls by GEO2R, Adjust $P<0.05$ and $|\log_2 \text{fold-times change} (|\log_2 \text{FC}|) > 1$, the log$_2$FC > 1 were defined as up-regulated genes and log$_2$FC < -1 as the down-regulated.

**GO and KEGG pathway enrichment analysis of DEGs**

After the DEGs were screened, GO and KEGG pathway analyses were further performed. GO analysis mainly included BP, CC, and MF. Screened DEGs were uploaded to Metascape gene annotation and resource analysis (https://metascape.org/gp/index.html#/main/step1). Fisher’s exact probability method was adopted to calculate $P$ values, and the screening conditions were all $P<0.05$. The $P$ values were ranked from small to large, and the first 20 GO functional families and KEGG signaling pathways were sorted out. Finally, the network diagram was drawn according to Cytoscape visual analysis.

**PPI network of DEGs and screening of key genes**

The potential relationships among DEGs were predicted by STRING (hops://www.string-db.org). High confidence > 0.7 was applied as the threshold to obtain the PPI map, and the data into Cytoscape (version 3.7.1). The module analysis of PPI was performed through the plugin MCODE to screen the PPI modules with a strong correlation.

**Results**

**Analysis of screening results**

GEO2R was employed to analyze the DEGs between AD patients and controls (Normal) in the GSE120721 dataset. Compared with controls, AD patients presented

![Volcano plot of DEGs](image-url)

**Figure 1.** Volcano plot of DEGs. (A: AD VS normal in GSE120721; B: L VS N in GSE120721; C: AD VS Normal in GSE32924; D: L VS N in GSE32924) (upregulated in red, downregulated in blue).

---

**Table 1.** Inclusion criteria for gene microarray datasets.

<table>
<thead>
<tr>
<th>No.</th>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Having microarray chip technology</td>
</tr>
<tr>
<td>2</td>
<td>Having comparison of microarray expression data in skin lesions of normal controls, patients with lesions and patients without lesions</td>
</tr>
<tr>
<td>3</td>
<td>The total sample size in the microarray data set should be greater than 15</td>
</tr>
</tbody>
</table>
991 up-regulation and 1,064 down-regulation genes. The DEGs between diseased (L) and non-diseased (N) AD patients were analyzed, with 567 up-regulation and 288 down-regulation genes in non-diseased AD patients in contrast with diseased patients. GSE32924 dataset had 799 up-regulation and 971 down-regulation genes in AD patients in contrast with controls. AD patients presented 11 up-regulation and 23 down-regulation genes in non-diseased in contrast with diseased patients (Figure 1).

The statistical results of DEGs in different lesion conditions in AD patients are illustrated in Figure 2. 78 DEGs in the GSE120721 dataset, and 9 DEGs in three groups in the GSE32924 dataset.

After the integration of DEGs, the statistical results of DEGs common to the two AD microarray expression datasets are illustrated in Figure 3. 80 co-expressed up-regulated DEGs and 75 co-expressed down-regulated were obtained, respectively.

**GO and KEGG pathway enrichment analysis of co-expressed DEGs**

GO analysis (Table 2 and Figure 4) suggested that the main enrichment function of DEGs was the regulation of response to biological stimulation, cell division, cellular macromolecular catabolic process, and the response of cells to stress. KEGG analysis revealed that DEGs were enriched in 8 pathways, containing the signal transduction of Rho GTPases, Miro GTPases, and RHOBTB3, cytokine signaling in the immune system, and receptor tyrosine kinase signaling. The Q value and RHOBTB3 were the smallest, and the number of DEGs enriched was the largest in Rho GTPases, and Miro GTPases, with a total of 106 genes.

**PPI map of DEGs and screening of key genes**

The candidate genes to imported into STRING 11.0 to

<table>
<thead>
<tr>
<th>GO Category</th>
<th>Description</th>
<th>Count</th>
<th>%</th>
<th>Log10(P)</th>
<th>Log10(q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reactome Gene Sets</td>
<td>Signaling by Rho GTPases, Miro GTPases and RHOBTB3</td>
<td>106</td>
<td>6.18</td>
<td>-18.59</td>
<td>-14.40</td>
</tr>
<tr>
<td>Reactome Gene Sets</td>
<td>Cytokine Signaling in Immune System</td>
<td>106</td>
<td>6.18</td>
<td>-18.28</td>
<td>-14.40</td>
</tr>
<tr>
<td>GO BP</td>
<td>regulation of response to biotic stimulus</td>
<td>78</td>
<td>4.55</td>
<td>-17.26</td>
<td>-13.51</td>
</tr>
<tr>
<td>Reactome Gene Sets</td>
<td>SARS-CoV Infections</td>
<td>71</td>
<td>4.14</td>
<td>-16.18</td>
<td>-12.53</td>
</tr>
<tr>
<td>GO BP</td>
<td>cell division</td>
<td>77</td>
<td>4.49</td>
<td>-14.38</td>
<td>-10.88</td>
</tr>
<tr>
<td>GO BP</td>
<td>regulation of cellular response to stress</td>
<td>96</td>
<td>5.59</td>
<td>-13.96</td>
<td>-10.69</td>
</tr>
<tr>
<td>GO BP</td>
<td>innate immune response</td>
<td>97</td>
<td>5.65</td>
<td>-13.68</td>
<td>-10.50</td>
</tr>
<tr>
<td>Reactome Gene Sets</td>
<td>Cellular responses to stress</td>
<td>92</td>
<td>5.36</td>
<td>-13.54</td>
<td>-10.40</td>
</tr>
<tr>
<td>Reactome Gene Sets</td>
<td>Vesicle-mediated transport</td>
<td>89</td>
<td>5.19</td>
<td>-12.98</td>
<td>-9.89</td>
</tr>
<tr>
<td>GO BP</td>
<td>regulation of cell-cell adhesion</td>
<td>72</td>
<td>4.20</td>
<td>-12.80</td>
<td>-9.73</td>
</tr>
<tr>
<td>GO BP</td>
<td>organelle localization</td>
<td>74</td>
<td>4.31</td>
<td>-12.51</td>
<td>-9.52</td>
</tr>
<tr>
<td>GO BP</td>
<td>regulation of cell cycle process</td>
<td>93</td>
<td>5.42</td>
<td>-11.99</td>
<td>-9.06</td>
</tr>
<tr>
<td>Reactome Gene Sets</td>
<td>Signaling by Receptor Tyrosine Kinases</td>
<td>73</td>
<td>4.25</td>
<td>-11.85</td>
<td>-8.95</td>
</tr>
<tr>
<td>Reactome Gene Sets</td>
<td>Cellular responses to stress</td>
<td>96</td>
<td>5.59</td>
<td>-11.77</td>
<td>-8.90</td>
</tr>
<tr>
<td>GO BP</td>
<td>post-Golgi vesicle-mediated transport</td>
<td>26</td>
<td>1.52</td>
<td>-11.05</td>
<td>-8.27</td>
</tr>
<tr>
<td>GO BP</td>
<td>intracellular protein transport</td>
<td>86</td>
<td>5.01</td>
<td>-10.96</td>
<td>-8.22</td>
</tr>
<tr>
<td>Reactome Gene Sets</td>
<td>Hemostasis</td>
<td>79</td>
<td>4.60</td>
<td>-10.74</td>
<td>-8.03</td>
</tr>
<tr>
<td>GO BP</td>
<td>positive regulation of hydrolase activity</td>
<td>74</td>
<td>4.31</td>
<td>-10.55</td>
<td>-7.87</td>
</tr>
</tbody>
</table>
obtain PPI, and the generated files were imported into Cytoscape to generate a PPI map (27 free targets). Each node of PPI represented a different DEG. The results of Cytoscape and MCODE clustering were compared and confirmed with each other, and their intersection genes were adopted as hub DEGs. 16 hub DEGs were obtained, which were potential key genes leading to the pathogenesis of AD. They were: FGD2, SYNE2, CC2D1A, PCDHB14, SCFD1, EIF3M, TAMM41, CSTF3, SPOP, ZNF623, CITED2, KPNB1, VPS37C, WRNIP1, YARS, and AGL (Figures 5 and 6).

Discussion

Bioinformatics-based screening methods were investigated to analyze the DEGs of AD and the pathogenesis. By analyzing two public microarray datasets, DEGs were obtained between AD patients and controls, as well as between diseased and non-diseased patients. In addition, the DEGs shared by the two datasets were integrated together for GO and KEGG pathway enrichment analysis, as well as the construction of PPI, and 16 hub DEGs were found as potential key genes.

DEGs between AD patients and controls and between diseased patients and non-diseased patients

GEO2R analysis suggested that there were many DEGs between AD patients and controls, as well as between diseased and non-diseased patients, indicating that abnormal expression regulation of genes is an important influencing factor in the pathogenesis of AD. It is worth noting that the number of DEGs between diseased and non-diseased patients in the GSE120721 dataset was clearly less than that between AD patients and controls, which may be because the gene expression difference between diseased and non-diseased patients in AD patients is not obvious enough, resulting in a small number of DEGs.

Another related study adopted gene chip technology to detect DEGs between AD patients and healthy controls in the Chinese population (16). 1,337 DEGs were identified, and 754 genes were up-regulated and 583 were down-regulated. GO and KEGG analysis showed that these DEGs were mainly involved in BP in immune response, inflammation regulation, and cell cycle. Similar to the results, they also found that the dysregulation of many immune system-related pathways and key genes has a major role in the pathogenesis of AD (17).

GO and KEGG pathway enrichment analysis

GO and KEGG analysis of the co-expressed DEGs indicated that these genes were mainly involved in the regulation of response to biological stimulation, cell division, cellular macromolecular catabolic process, and regulating cellular response to stress. In addition, KEGG analysis revealed that DEGs were mainly enriched in 8 pathways, including Rho GTPases, Miro GTPases, and RHOBTB3 signaling, cytokine signaling in the immune system, and receptor tyrosine kinase signaling. These pathways are important cues in the pathogenesis of AD.

KEGG analysis displayed that DEGs were clearly enriched in Rho GTPases, Miro GTPases, RHOBTB3, immune system cytokine, receptor tyrosine kinase, and other signaling pathways. Among them, Rho GTPases, Miro GTPases, and RHOBTB3 had the smallest q values and the most significant enrichment of DEGs, with 106 genes. These pathways have been implicated in the pathogenesis of AD in previous reports. It has been found that Rho GTPase signaling plays a key role in regulating actin cytoskeleton organization and cell adhesion, both of which are involved in skin barrier function and immune responses in AD (18,19). Similarly, the Miro GTPases pathway is involved in mitochondrial transport and dynamics, which are essential for cellular metabolism and have been shown to be impaired in AD (20). Other studies have analyzed gene differences between AD patients and controls through GEO data sets and found genes consistent with this result (21). They found an enrichment of the Rho
GTPases signaling pathway, which is also an enriched pathway in the present finding. In addition, they also found that many genes were enriched in cytokine signaling in the immune system, which is also consistent with the present results (22).

### PPI analysis and key gene screening

To further explore potential interactions among DEGs, a PPI was constructed using the STRING database and then visualization was carried out adopting Cytoscape. The resulting PPI consisted of 27 target genes, which were further analyzed by adopting the MCODE clustering algorithm to identify hub genes. 16 hub genes having FGD2, SYNE2, CC2D1A, PCDH14, SCFD1, EIF3M, TAMM41, CSTF3, SPOP, ZNF623, CITED2, KPNB1, VPS37C, WRNIP1, YARS, and AGL were identified. These genes take part in several BP, likely immune response, cell proliferation, differentiation, and DNA repair. Importantly, several of these genes are consistent with previous studies mentioning AD pathogenesis, including FGD2, SYNE2, CC2D1A, and SPOP. In addition, the molecular mechanism of AD pathogenesis was explored and many genes were found to be identical to the present results (23). For example, they also found some genes enriched in signaling in response to biological stimuli and in the immune system. These studies also mentioned some genes such as SYNE2, CITED2, KPNB1, and YARS, which are also hub DEGs in this work (24).

In summary, several enriched pathways potentially involved in AD pathogenesis were identified, having genes involved in immune response, cell proliferation and differentiation, and DNA repair. Identification of hub genes using PPI analysis further provides potential therapeutic targets for AD treatment. However, there is a need to further validate the role of these genes and pathways in AD pathogenesis and explore their potential as therapeutic targets. In conclusion, DEGs, screening and enrichment analysis based on bioinformatics can reveal the key BP and molecular mechanism in the pathogenesis of AD, which provides an important basis for future research. However, it should be noted that due to the influence of factors such as individual differences, sample size, and source, the enrichment results of different DEGs and pathways may differ, so it is necessary to confirm each other and further verify and improve these results by multiple independent studies.

By analyzing the two AD chip expression datasets in the GEO database, 80 co-expressed up-regulated DEGs and 75 co-expressed down-regulated DEGs were screened. GO and KEGG analysis revealed that these DEGs were mainly involved in biological stimulation response, cell division, cell metabolism, and other functions, and were enriched in multiple signaling pathways. For example, Rho GTPases, Miro GTPases, and RHOBTBB signal transduction, cytokine signaling in the immune system, and receptor tyrosine kinase signaling. By importing the candidate genes into the STRING 11.0 database and generating PPI, 16 hub DEGs were screened out, which were potential key genes leading to the pathogenesis of AD, such as FGD2, SYNE2, CC2D1A, PCDHB14. Therefore, it reveals the potential mechanism of AD and provides an important reference for AD.

### References


