SPRR2C, DEFB4A, WIF1, CRY2, and KRT19 are correlated with the development of atopic eczema

Yiwu Yu  
Tongji University Affiliated Shanghai Tongji Hospital

Yufei Li  
Tongji University Affiliated Shanghai Tongji Hospital

Miao Jiang  
Tongji University Affiliated Shanghai Tongji Hospital

Jingjun Zhao (✉ zhaomyco@163.com)  
Tongji Hospital of Tongji Medical College of Tongji University of medicine

Research

Keywords: atopic eczema, differentially expressed genes, weighted correlation network, functional and pathway enrichment analyses

DOI: https://doi.org/10.21203/rs.3.rs-57989/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

Atopic eczema (AE) is a chronic relapsing inflammatory skin disease. The objective of this study was to identify key genes related to the development of AE.

Methods

The GSE6012 dataset was obtained from the Gene Expression Omnibus (GEO) database. The limma package was used to analyze differentially expressed genes (DEGs). Then, the weighted gene co-expression network analysis (WGCNA) package was utilized to generate weighted correlation networks of up- and downregulated genes. Additionally, the WGCNA package was used for enrichment analyses to explore the underlying functions of DEGs in modules (weighted correlation sub-networks) significantly associated with AE.

Results

A total of 515 DEGs were identified between lesional and non-lesional skin samples. For the upregulated genes, the blue module was found to have a significant positive correlation with AE. Importantly, small proline-rich protein 2C (SPRR2C) and defensin, beta 4A (DEFB4A) exhibited higher |log fold change (FC)| values and were the key nodes of the network. Moreover, KEGG pathway analysis revealed that the upregulated genes in the blue module were primarily involved in cytokine-cytokine receptor interaction. Additionally, for the downregulated genes, the brown module was found to have a significant positive correlation with AE. Further, WNT inhibitory factor 1 (WIF1), cryptochrome 2 (CRY2), and keratin 19 (KRT19) had higher |log FC| values and were key nodes of the network.

Conclusion

SPRR2C, DEFB4A, WIF1, CRY2, KRT19 and cytokine-cytokine receptor interaction might be correlated with the development of AE.

Background

Atopic eczema (AE, also known as eczema or atopic dermatitis) is a non-contagious, inflammatory, relapsing, and itchy skin disorder [1, 2]. AE patients often have scaly and dry skin spanning almost their entire body, along with intensely itchy red lesions that are at high risk of viral, bacterial, and fungal colonization [3–5]. AE affects 2–10% of adults and 15–30% of children in developed countries, and the rate has approximately tripled in the United States over the past 30–40 years [6–8].

Recently, several studies have investigated the mechanisms of AE development. FLG null alleles predispose patients to a type of eczema that persists from early infancy to adulthood and can act as an indicator of poor prognosis in AE patients [9]. The chromosomal region containing the low-affinity Fc
receptor for the IgE (FCER2) gene has a regulatory function in atopic disorders [10]. Mutations of nucleotide-binding oligomerization domain protein 1 (NOD1) is closely related with atopy susceptibility [11]. The C-1237T promoter polymorphism of toll-like receptor 9 (TLR9) may play a role in AE susceptibility, particularly in patients with an intrinsic AE variant [12]. As the predominant aquaporin in human skin, aquaporin 3 (AQP3) expression is upregulated and has altered cellular distribution in eczema, which may result in water loss [13]. The levels of reduced soluble cluster of differentiation 14 (sCD14) in the fetal and neonatal gastrointestinal tract are related to the development of eczema or atopy, and thus sCD14 may be used in disease treatment [14]. However, despite the above findings, the detailed mechanisms underlying AE remain to be elucidated.

Using the data of Mobini et al. [15], we further screened differentially expressed genes (DEGs) in AE. Additionally, weighted correlation networks were constructed separately for the up- and downregulated genes. The potential functions of DEGs in modules (weighted correlation sub-networks) that were significantly correlated with AE were analyzed by Gene Ontology (GO) and pathway enrichment analyses. We expected to identify genes associated with AE development and provide novel therapeutic targets for AE. This study may contribute new insights into the changes in gene expression in AE, potentially revealing the underlying mechanisms of AE.

Materials And Methods

1.1 Microarray data

The GSE6012 dataset was obtained from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. This dataset was generated using a GPL96 [HG-U133A] Affymetrix Human Genome U133A Array. The GSE6012 dataset included 10 skin biopsies from patients with AE and 10 skin biopsies from healthy controls.

1.2 Data preprocessing and differentially gene expression analysis

After the GSE6012 dataset was downloaded, the microarray data were preprocessed according to the following steps. First, probe names were converted to gene names. Next, for genes mapped with multiple probes, the final gene expression value was determined as the average value of each probe. Expression values were log2 transformed and normalized by the preprocessCore package [16] in R. The distribution of gene expression values before and after normalization are displayed in a box plot. The limma package [17] in R was utilized to analyze the DEGs between lesional and non-lesional skin samples with the cutoff being adjusted p-value < 0.05 and |log fold change (FC)| > 1.

1.3 Weighted correlation network construction

The weighted gene co-expression network analysis (WGCNA) package [18] in R was utilized to generate weighted correlation networks for the up- and downregulated genes. Briefly, gene clustering was performed using the expression matrix of the DEGs. Next, DEGs were further screened by removing
outliers. The criteria was adjusted several times to include no more than 3,600 DEGs. A correlation coefficient of no less than 0.8 was set as the weighting coefficient. Finally, the correlation matrix was converted to a topological matrix.

Hierarchical clustering was performed separately for the up- and downregulated genes using a hybrid dynamic shear tree method, and different branches of the clustering tree represented different gene modules. The minimum number of genes involved in each gene module was set to 30. Subsequently, the feature vector was calculated for each module and cluster analysis was performed on the modules. Modules that clustered closely were merged into new modules. Cytoscape [19] was utilized to visualize the weighted correlation network. Correlation analysis between gene modules and AE was carried out using the correlation coefficient method and network significance method, respectively.

1.4 Functional and pathway enrichment analyses

Using the WGCNA package in R, GO [20], enrichment analysis was performed on the DEGs in modules significantly correlated with AE. Additionally, the culsterprofiler package in R was used to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) [21] pathway enrichment analysis separately for up- and downregulated genes in the modules most significantly correlated with AE. GO and KEGG pathway enrichment analyses were performed on the top 30 DEGs in key modules. The cut-off criterion was p-value < 0.05.

Discussion

In this study, a total 515 DEGs between lesional and non-lesional skin samples were screened, including 286 up- and 229 downregulated genes. The enriched functions for the up- and downregulated genes were primarily in the BP category.

Being an allergen-induced gene in experimental allergic responses, small proline-rich protein 2 (SPRR2) may be associated with allergic inflammation [22]. SPR1 and SPR2 in the epidermis, as well as SPR3 in cultured keratinocytes, may be epidermal cell envelope (CE) components [23, 24]. Some members of the SPRR-family are overexpressed during ageing, thereby reducing the skin's barrier function against hostile attacks from the environment [25]. We found that SPRR2C was among the top 30 upregulated DEGs in the blue module and was significantly correlated (correlation coefficient = 0.94) with AE, indicating that the expression levels of SPRR2C may be related to AE.

As a peoptide antibiotic, human b-defensin-2 (hBD-2, also known as DEFB4A) can protect human skin with psoriasis from infection by abnormal expression in response to invasion of microorganisms [26]. Overexpression of hBD-2, which can be induced by Staphylococcus aureus, can cause persistent eczematous skin lesions in patients with AE [27]. Expression of hBD-2 can be induced by injury, inflammatory stimuli, and bacteria on the skin of AE patients [28]. In addition, serum hBD-2 levels may be enhanced by oncostatin M and interleukin-22 (IL-22) through transcription 3 (STAT3) in keratinocytes, and this may also function as a biomarker of skin inflammation [29]. Additionally, the correlation between
high DEFB4 expression and psoriasis risk indicates that hBD2 may play a role in the skin's inflammatory response [30]. We also found that upregulated DEFB4A was among the top 30 DEGs in the blue module significantly correlated (correlation coefficient = 0.94) with AE, indicating that DEFB4A might play a role in AE.

Abnormal circadian clockwork often induces bipolar disorders and depression in patients. As a circadian gene, CRY2 is involved in regulating the evening oscillator [31]. CRY2 exists in the epidermis, plays an important role in maintaining the epidermal clock, and cannot be replaced by external light [32]. The variable expression of KRT15 and KRT19 in oral squamous cell carcinoma (OSCC) and squamous intraepithelial neoplasm (SIN) results in their divergent biological behaviors and roles in pathogenesis, indicating that they may be used as markers to classify OSCC and SIN [33]. KRT19 may also function as a specific epithelial marker [34]. Expression of KRT19 in the skin may be an additional characterization of skin stem cells under pathological and normal conditions [35]. In our study, we found that downregulated CRY2 and KRT19 were among the top 30 DEGs in the brown module significantly correlated (correlation coefficient = 0.97) with AE. These might indicate that the expression levels of CRY2 and KRT19 are related to AE.

Downregulation of WIF1 is implicated in melasma development by upregulating the canonical or noncanonical Wnt signaling pathway [36]. Although Wnt signaling regulates skin pigmentation, WIF1 is expressed not only in melanocytes, but also in keratinocytes [37, 38] and fibroblasts [39]. WIF1 expression is upregulated in interfollicular keratinocyte stem cells (KSCs), which is of great interest given the increased levels of Wnt signaling in psoriasis, wound healing, and basal cell carcinomas [40, 41]. WIF1 can function as a marker of interfollicular KSCs and can inhibit cell cycle progression in human keratinocytes, even under the activation of Wnt signals (Wnt3A) [42]. We also found that downregulated WIF1 was among the top 30 DEGs in the brown module significantly correlated (correlation coefficient = 0.97) with AE. This indicated that WIF1 might be associated with AE.

In this study, KEGG pathway analysis revealed that the upregulated genes in the blue module were primarily involved in the cytokine-cytokine receptor interaction. Cytokine-cytokine receptor interaction is associated with the progression of skin-related diseases by regulating the proliferation of skin-derived precursors (SKPs) and SKP differentiation [43]. In addition, cytokine-cytokine receptor interaction has previously been demonstrated to play an important role in the progression of AE [44]. Thus, we inferred that cytokine-cytokine receptor interaction might be closely correlated with the AE progression.

In the present study, a weighted correlation network analysis was utilized to identify key genes involved in AE, providing a basis for further study of AE. However, relevant experiments should be conducted to verify the numerous candidate genes and signaling pathways identified in this study. In addition, in depth research is required to elucidate the specific mechanisms of action in AE.

In conclusion, we performed a comprehensive bioinformatics analysis of genes which may be involved in AE. SPRR2C, DEFB4A, WIF1, CRY2, KRT19, and cytokine-cytokine receptor interaction might play a role in AE.
Abbreviations

AE: Atopic eczema; GEO: Gene Expression Omnibus; DEGs: differentially expressed genes

WGCNA: weighted gene co-expression network analysis; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

Declarations

Availability of data and materials

The datasets analyzed during the current study are available in the GEO database (https://www.ncbi.nlm.nih.gov/geo).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Funding

This research was supported by grants from the National Natural Science Foundation of China (81874240) to Jingjun Zhao and National Natural Science Foundation of China (81803126) to Miao Jiang.

Author Contributions

Prof. Jingjun Zhao and Miao Jiang conceived and designed the research study and wrote the paper. Yiwu Yu and Yufei Li performed the experiments, analyzed the data and wrote the paper. All authors read and approved the manuscript.

Acknowledgments

Not applicable
References


**Tables**

**Table 1.** The result of correlation coefficient test between modules screened for the significantly DEGs and atopic eczema.

(A) The result of correlation coefficient test between modules screened for the significantly up-regulated genes and atopic eczema.

<table>
<thead>
<tr>
<th>Module</th>
<th>ME-black</th>
<th>ME-green</th>
<th>ME-blue</th>
<th>ME-red</th>
<th>ME-grey</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.69</td>
<td>0.74</td>
<td>0.94</td>
<td>0.79</td>
<td>0.44</td>
</tr>
<tr>
<td>P-value</td>
<td>0.000746531</td>
<td>0.000169191</td>
<td>5.76E-10</td>
<td>3.48E-05</td>
<td>0.05152951</td>
</tr>
</tbody>
</table>
(B) The result of correlation coefficient test between modules screened for the significantly down-regulated genes and atopic eczema.

<table>
<thead>
<tr>
<th>Module</th>
<th>ME-blue</th>
<th>ME-brown</th>
<th>ME-yellow</th>
<th>ME-grey</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.72</td>
<td>0.97</td>
<td>0.73</td>
<td>0.52</td>
</tr>
<tr>
<td>P-value</td>
<td>0.000368267</td>
<td>7.60E-13</td>
<td>0.000251289</td>
<td>0.01867507</td>
</tr>
</tbody>
</table>

Figures
Figure 1

Normalized expression value data. The box in the black line indicates the median of each data set, which determines the degree of standardization through the distribution of the data. Following normalization, the black line in the box is almost in the same straight line, indicating a good degree of standardization.

Figure 2

Heatmap and volcano plot of differentially expressed genes (DEGs) between lesional and non-lesional skin samples. a. Heatmap of DEGs; b. Volcano plot of DEGs.
Figure 3

Construction of weighted correlation sub-networks for upregulated genes. a. The gene clustering tree constructed for upregulated genes; b. The selection of the weighting coefficient (the horizontal axis represents the soft threshold power and the vertical axis represents the square of the correlation coefficient between log2k and log2p (k). The blue line indicates where the correlation coefficient is 0.8, and the corresponding soft threshold power is 7); c. Clustering result of weighted correlation sub-networks (different modules are indicated by colors underneath the dendrogram); d. Mean values of gene significances for differentially expressed genes (DEGs) in the modules; e. Heatmap of the top 30 genes from the module significantly correlated with atopic eczema.
Figure 4

Constructing weighted correlation sub-networks for downregulated genes. a. The gene clustering tree constructed for downregulated genes; b. The selection of the weighting coefficient (the horizontal axis represents the soft threshold power and the vertical axis represents the square of the correlation coefficient between log2k and log2p (k). The blue line indicates where the correlation coefficient is 0.8, and the corresponding soft threshold power is 5); c. Clustering result of weighted correlation sub-networks (different modules are indicated by colors underneath the dendrogram); d. Mean values of gene significances for differentially expressed genes (DEGs) in the modules; e. Heatmap of the top 30 genes from the module significantly correlated with atopic eczema.
Figure 5

Weighted correlation network for the top 30 differentially expressed genes (DEGs) from the blue module. The color depth of nodes indicate the fold change values of corresponding genes. The thickness of edges represent the co-expression coefficients of genes.
Figure 6

Weighted correlation network for the top 30 differentially expressed genes (DEGs) screened from the brown module. The color depth of nodes indicate the fold change values of corresponding genes. The thickness of edges represent the co-expression coefficients of genes.