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

Y.-W. YU, Y.-F. LI, M. JIANG, J.-J. ZHAO

Department of Dermatology, Tongji Hospital, Tongji University School of Medicine, Shanghai, China

Abstract. - OBJECTIVE: Atopic eczema (AE) is a chronic relapsing inflammatory skin disease. This study aims to identify key genes related to the development of AE.

MATERIALS AND 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 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.

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

Key Words: Atopic eczema, Differentially expressed genes, Weighted correlation network, Functional and pathway enrichment analyses.

Introduction

Atopic eczema (AE, also known as eczema or atopic dermatitis) is a non-contagious, inflammatory, relapsing, and itchy skin disorder. 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. 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.

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. The chromosomal region containing the low-affinity Fc receptor for the IgE (FCER2) gene has a regulatory function in atopic disorders. Mutations of nucleotide-binding oligomerization domain protein 1 (NOD1) are closely related with atopy susceptibility. 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. 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. 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. However, despite the above findings, the detailed mechanisms underlying AE remain to be elucidated.
SPRR2C, DEFB4A, WIF1, CRY2, and KRT19 are correlated with the development of atopic eczema

Using the data of Mobini et al., 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**

**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.

**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 preprocesCore package in R. The distributions of gene expression values before and after normalization are displayed in a box plot. The limma package 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.

**Weighted Correlation Network Construction**

The weighted gene co-expression network analysis (WGCNA) package 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 were 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 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.

**Functional and Pathway Enrichment Analyses**

Using the WGCNA package in R, GO 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) 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.

**Results**

**Data Preprocessing and Identification of DEGs**

Following data preprocessing, 12,927 genes were mapped to the probes. Gene expression profiles prior to and following normalization are shown in Figure 1. The results indicate that the median probe intensity was similar across all conditions, demonstrating the consistency of the technical quality of the data following normalization.

A total of 515 significantly DEGs, between lesional and non-lesional skin samples, including 286 upregulated and 229 downregulated genes, were screened with the criteria of adjusted p-value < 0.05 and |log FC| > 1. The heatmap and volcano plot of DEGs are shown in Figure 2. The two groups of samples can be significantly separated, indicating the reliability of the analysis.
**Weighted Correlation Network Construction**

According to the gene clustering tree (Figure 3a), a total of 256 upregulated genes were used to construct a weighted correlation network. The weighting coefficient was set to 7 (Figure 3b). A total of five modules (namely the red module, blue module, grey module, black module, and green module) were screened for the upregulated genes (Figure 3c). The correlation coefficient method (Table IA) revealed that the blue module had a significant positive correlation (correlation coefficient = 0.94) with AE. The network significance method (Figure 3d) revealed that the blue module had the highest Module Significance (MS) value (that is, the average value of gene significance, MS value = 0.62) when compared with other modules. This was in accordance with the result of the correlation coefficient method.

Based on the gene clustering tree (Figure 4a), a total of 225 downregulated genes were used to construct a weighted correlation network. The weighting coefficient was set to 5 (Figure 4b). A total of four modules (namely the brown module, blue module, grey module and yellow module) were screened for the significantly downregulat-

---

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

1439

ed genes (Figure 4c). The correlation coefficient method (Table IB) revealed that the brown module had a significant positive correlation (correlation coefficient = 0.97) with AE. The network significance method (Figure 4d) revealed that the brown module had the highest MS value (MS value = 0.61) compared with other modules. This was in accordance with the result of the correlation coefficient method.

**Functional and Pathway Enrichment Analyses**

For the upregulated genes in the blue module, enriched GO functions included 223 terms in the biological process (BP) category, 5 terms in the cellular component (CC) category, and 19 terms in the molecular function (MF) category. Most of the enriched functions were in the BP category. There were five enriched pathways in the upregulated genes in the blue module, including rheumatoid arthritis ($p = 0$), cytokine-cytokine receptor interaction ($p = 0$), staphylococcus aureus infection ($p = 0$), chemokine signaling pathway ($p = 0$), and Toxoplasmosis ($p = 0.01$) (Table IIB).

Based on the connectivity of the blue module (Figure 3e), the top 30 DEGs were used to construct a weighted correlation network. The weighted correlation network had 30 nodes and 435 interactions (Figure 5). Importantly, lac-

---

**Table I.** The result of correlation coefficient test between modules screened for the significantly DEGs 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>Atopic eczema</td>
<td>coefficient</td>
<td>0.69</td>
<td>0.74</td>
<td>0.94</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.000746531</td>
<td>0.000169191</td>
<td>5.76E-10</td>
<td>3.48E-05</td>
</tr>
</tbody>
</table>

<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>Atopic eczema</td>
<td>coefficient</td>
<td>0.72</td>
<td>0.97</td>
<td>0.73</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.000368267</td>
<td>7.60E-13</td>
<td>0.000251289</td>
<td>0.01867507</td>
</tr>
</tbody>
</table>
totransferrin (LTF), small proline-rich protein 2C (SPRR2C) and defensin, beta 4A (DEFB4A) had higher |log (FC)| values and were the key nodes of the network.

The enriched GO functions for the top 30 DEGs of the blue module are listed in Table IIA. The enriched functions in the BP category included cellular process \((p = 0.006748448)\), single-organism cellular process \((p = 0.000188025)\), and metabolic process \((p = 0.003581022)\). The enriched functions in the MF category included catalytic activity \((p = 0.002039884)\) and binding \((p = 0.011042218)\). In addition, there were no enriched KEGG pathways for the top 30 DEGs of the blue module.

For the downregulated genes in the brown module, the enriched functions in the BP category included biological_process \((p = 2.25E-11)\), single-organism cellular process \((p = 9.94E-05)\), and single-multicellular organism process \((p = 7.06E-06)\). The enriched functions in the CC category included cellular_component \((p = 0.000364364)\) and extracellular region part \((p = 7.44E-05)\). The enriched functions in the MF category included molecular_function \((p = 1.33E-09)\) and binding \((p = 0.002472104)\) (Table IIIA).
SPRR2C, DEFB4A, WIF1, CRY2, and KRT19 are correlated with the development of atopic eczema

Table II. The enriched GO functions and KEGG pathways for the significantly up-regulated genes of the blue module.

<table>
<thead>
<tr>
<th>Term ID</th>
<th>Description</th>
<th>Gene number</th>
<th>Gene symbol</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>GO:0008150 biological_process</td>
<td>23</td>
<td>GALNT6, DSC2...</td>
<td>0.010350733</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0009987 cellular process</td>
<td>22</td>
<td>DSC2, TYMP</td>
<td>0.006748448</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0044763 single-organism cellular process</td>
<td>21</td>
<td>TYMP, EPHA1</td>
<td>0.000188025</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0044699 single-organism process</td>
<td>21</td>
<td>IRF7, ISG20</td>
<td>0.001240746</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0008152 metabolic process</td>
<td>19</td>
<td>EPHA1, IRF7</td>
<td>0.003581022</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0016043 cellular component organization</td>
<td>12</td>
<td>TYMP, EPHA1</td>
<td>0.003946095</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0032502 developmental process</td>
<td>12</td>
<td>IRF7, MMP12</td>
<td>0.005918045</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0070887 cellular response to chemical stimulus</td>
<td>7</td>
<td>IRF7, ISG20</td>
<td>0.008117276</td>
</tr>
<tr>
<td>MF</td>
<td>GO:0003824 catalytic activity</td>
<td>13</td>
<td>GALNT6, TYMP</td>
<td>0.002039884</td>
</tr>
<tr>
<td>MF</td>
<td>GO:0005488 binding</td>
<td>19</td>
<td>DSC2, TYMP</td>
<td>0.011042218</td>
</tr>
</tbody>
</table>

Table III. The enriched GO functions for the significantly down-regulated genes of the brown module.

<table>
<thead>
<tr>
<th>Term ID</th>
<th>Description</th>
<th>Gene number</th>
<th>Gene symbol</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>GO:0008150 biological_process</td>
<td>123</td>
<td>C1orf68, MPHOSPH6</td>
<td>2.5E-11</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0044699 single-organism process</td>
<td>88</td>
<td>SLCO1B1, ADIRF</td>
<td>9.9E-05</td>
</tr>
<tr>
<td>BP</td>
<td>GO:0044707 single-multicellular organism process</td>
<td>62</td>
<td>C1orf68, WIFI...</td>
<td>7.1E-08</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0005575 cellular_component</td>
<td>131</td>
<td>SLCO1B1, SCGB1D2</td>
<td>0.000364364</td>
</tr>
<tr>
<td>CC</td>
<td>GO:0005576 extracellular region</td>
<td>30</td>
<td>WIFI1, PRR4</td>
<td>0.000390111</td>
</tr>
<tr>
<td>CC</td>
<td>GO:004421 extracellular region part</td>
<td>21</td>
<td>CTGF, DEFA6</td>
<td>7.4E-05</td>
</tr>
<tr>
<td>MF</td>
<td>GO:0003674 molecular_function</td>
<td>123</td>
<td>ADIRF, POLI</td>
<td>1.3E-09</td>
</tr>
<tr>
<td>MF</td>
<td>GO:0005488 binding</td>
<td>96</td>
<td>CNN1, CRABP1</td>
<td>0.002472104</td>
</tr>
<tr>
<td>MF</td>
<td>GO:0005515 protein binding</td>
<td>66</td>
<td>POLI, CNN1</td>
<td>0.004562283</td>
</tr>
</tbody>
</table>
Based on the connectivity of the brown module (Figure 4e), the top 30 DEGs were used to construct a weighted correlation network. The weighted correlation network consisted of 30 nodes and 435 interactions (Figure 6). Further, WNT inhibitory factor 1 (WIF1), secretoglobin, family 1D, member 2 (SCGB1D2), cryptochrome 2 (CRY2), and keratin 19 (KRT19) had higher \(|\log FC|\) values and were the key nodes of the network.

The enriched GO functions for the top 30 DEGs of the brown module are listed in Table IIB. The enriched functions in the BP category included single-multicellular organism process \((p = 3.19E-05)\), multicellular organismal process \((p = 5.41E-05)\), and single-organism process \((p = 0.000161257)\). The enriched functions in the MF category included protein binding \((p = 0.001386242)\) and binding \((p = 0.011042218)\).

**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 \(\log k\) and \(\log p\) (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.
SPRR2C, DEFB4A, WIF1, CRY2, and KRT19 are correlated with the development of atopic eczema.

Figure 5. Weighted correlation network for the top 30 differentially expressed genes (DEGs) from the blue module. The color depth of nodes indicates the fold change values of corresponding genes. The thickness of edges represents 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 indicates the fold change values of corresponding genes. The thickness of edges represents the co-expression coefficients of genes.
Additionally, there were no enriched KEGG pathways for the top 30 DEGs of the brown module.

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 peptide 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 melanoma 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.

Conclusions

This paper presented 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.
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.

**Conflict of Interest**
The Authors declare that they have no conflict of interests.

**Acknowledgments**
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 Yufei Li conceived and designed the research study and wrote the paper. Yiwu Yu performed the experiments, analyzed the data and wrote the paper. All authors read and approved the manuscript.

**References**
16) Bolstad B. preprocessCore: A collection of preprocessing functions. R package version. 2013; 1.42.0


43) Zhou R, Mao Y, Xiong L, Li L. Integrated Transcriptome Analysis of microRNA and mRNA in Mouse Skin Derived Precursors (SKPs) and SKP Derived Fibroblast (SFBs) by RNA-Seq. Curr Genomics 2019; 20: 49-60.

44) Lim HS, Ha H, Shin HK, Jeong SJ. The Genome-Wide Expression Profile of Saussurea lappa Extract on House Dust Mite-Induced Atopic Dermatitis in Nc/Nga Mice. Mol Cells 2015; 38: 765-772.