Screening for key genes associated with atopic dermatitis with DNA microarrays

ZHONG-KUI ZHANG1, YONG YANG2, SHU-RONG BAI3, GUI-ZHEN ZHANG4, TAI-HUA LIU4, ZHOU ZHOU1, CHUN-MEI WANG1, LI-JUN TANG1, JUN WANG1 and SI-XIAN HE1

Departments of 1Dermatology, 2Burns, Plastic Surgery, 3Anesthesiology and 4Information, Chengdu Military General Hospital, Chengdu, Sichuan 610083, P.R. China

Received August 15, 2013; Accepted December 16, 2013

DOI: 10.3892/mmr.2014.1908

Abstract. The aim of the present study was to identify key genes associated with atopic dermatitis (AD) using microarray data and bioinformatic analyses. The dataset GSE6012, downloaded from the Gene Expression Omnibus (GEO) database, contains gene expression data from 10 AD skin samples and 10 healthy skin samples. Following data preprocessing, differentially expressed genes (DEGs) were identified using the limma package of the R project. Interaction networks were constructed comprising DEGs that showed a degree of node of >3, >5 and >10, using the Osprey software. Functional enrichment and pathway enrichment analysis of the network comprising all DEGs and of the network comprising DEGs with a high degree of node, were performed with the DAVID and WebGestalt toolkits, respectively. A total of 337 DEGs were identified. The functional enrichment analysis revealed that the list of DEGs was significantly enriched for proteins related to epidermis development (P=2.95E-07), including loricin (LOR), keratin 17 (KRT17), small proline-rich repeat proteins (SPRRs) and involucrin (IVL). The chemokine signaling pathway was the most significantly enriched pathway (P=0.0490978) in the network of all DEGs and in the network consisting of high degree-node DEGs (>10), which comprised the genes coding for chemokine receptor 7 (CCR7), chemokine ligand (CCL19), signal transducer and activator of transcription 1 (STAT1), and phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1). In conclusion, the list of AD-associated proteins identified in this study, including LOR, KRT17, SPRRs, IVL, CCR7, CCL19, PIK3R1 and STAT1 may prove useful for the development of methods to treat AD. From these proteins, PIK3R1 and KRT17 are novel and promising targets for AD therapy.

Introduction

Atopic dermatitis (AD), also known as atopic eczema (AE), is a chronic relapsing inflammatory skin disease. It has been classified into 3 sequential phases: infantile, childhood and adult, each with characteristic physical findings (1). Approximately 70% of patients have a family history of allergies (2). The disease has an important impact not only on patients’ physical and mental health, but also on the economy of society and affected families (3). In recent years, the incidence rate of AD has increased to almost 20% (4).

Various internal factors (i.e., gene mutations) and external factors (i.e., infection and allergens) influence immune cells and trigger abnormal immune response, which can eventually lead to AD (5). Defects in the innate immune system are also believed to contribute to pathogenesis of AD (6,7). As the pathophysiologic complexity of this disease is being elucidated, it becomes clear that the development of AD is triggered by complex interactions between genetic and environmental factors, along with interactions occurring within a dense network of cytokines and chemokines; however, a unifying concept for the pathogenesis of AD has not been established (8). Basic therapeutic treatment of AD includes optimal skin care, addressing skin barrier defects, avoidance of specific and non-specific factors that trigger AD, as well as local treatment, including wet-wrap or antimicrobial therapy and administration of glucocorticosteroids and other compounds (9). However, these therapies have a relatively slow onset of action and cannot systematically eradicate the symptoms interfering with daily life of the patients. Therefore, understanding the immunological basis of AD and developing effective immune modulators is expected to be the most effective strategy for treatment of this disease (10).

Omics technologies including proteomics (11) and microarrays (12) are powerful tools to identify the underlying molecular mechanisms of disease and potential biomarkers. Park et al identified 31 atopic dermatitis-related candidate proteins from a proteomic analysis (13). Microarray technology is also used to analyze the differences between disease and healthy samples (14). For instance, the study by Nomura et al
(15), used DNA microarray analysis to compare the complex gene expression pattern of skin lesions from AD and psoriasis.

In the present study, gene expression profiles of AD skin were compared to those of healthy skin to identify differentially expressed genes (DEGs). Gene network construction combined with function and pathway enrichment analysis were performed to investigate the functions and potential roles of the identified DEGs in the pathogenesis of AD.

Materials and methods

Microarray data. The microarray dataset GSE6012 was downloaded from the Gene Expression Omnibus database (GEO; http://www.ncbi.nlm.nih.gov/geo/) which contains data from 10 human AD samples (aged 25-50 years) and 10 healthy skin samples (aged 21-50 years). These data were generated from a previous study (16), and had been used for meta-analyses by other research groups (17). Briefly, the data were derived from the profiling of patients diagnosed with AD based on the criteria of Williams et al (18), which did not undergo local or systemic treatment with glucocorticoids or tacrolimus for the last 2 weeks prior to the study. The control groups of the study comprised healthy, symptom-free individuals, gender- and age-matched to the patient group, with normal levels of IgE, and no history of AE or any other atopic or chronic disease. Individuals in both the patient and the control group were of Caucasian origin. Two punch biopsies (4 mm diameter) were collected under local anaesthesia to obtain lesional skin samples from each patient and skin samples from the matching location of each control individual. Hybridizations were performed on the GPL96 Affymetrix Human Genome U133A (HG-U133A) array. Annotation files for human genes were also simultaneously gathered when the original expression data were collected from the GEO database.

Data treatment and identification of DEGs. The original CEL-format file was converted into an expression profile format with the open-source package Affy (19,20), available from the R project website (http://www.r-project.org/). The data were normalized with the median method. The expression profiles of AD and healthy samples were then analyzed with the linear models for microarray data (LIMMA) package available in Bioconductor and implemented in R language (21). Multiple testing correction was performed on the P-values using the Benjamini-Hochberg (BH) method (22). False discovery rate (FDR) and fold change (FC) were used as the criteria to identify DEGs, with cut-offs being FDR <0.05 and log fold change (FC) ≥1.

Construction of interaction networks for DEGs. In order to detect potential interactions among the DEGs, the previously described software Osprey (23) was chosen to retrieve known interaction relationships and construct interaction networks. Osprey is a suitable tool for this analysis, since it integrates information from the Biomolecular Interaction Network Database (BIND) (24) and the Global Resource Information Database (GRID) (25), while also including data on 50,000 interactions between proteins and nucleic acids. Osprey is used to build color-coded graphical representations of the interaction data. Networks built in Osprey can be saved as tab-delimited text files for future manipulation, or exported as Joint Photographic Experts Group (JPEG), portable network graphics (PNG), or scalable vector graphic (SVG) files (23).

Network screening for high degree-node DEGs. The degree of nodes represents the number of edges (here, other genes) linked to a node (a given gene) (26). The number of gene interactions correlates with the importance of the gene in the entire network and this hypothesis has been experimentally validated in the yeast Saccharomyces cerevisiae (27). In the present study, low-degree nodes were gradually removed, which led to a final network comprising DEGs with a degree of node ≥10.

Functional enrichment analysis. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) (28) is widely used in functional enrichment analysis of DEGs. In this study, DAVID was used to perform the functional enrichment analysis for the genes comprised in the network of all DEGs and those comprised in the network of DEGs with a high degree of node (≥10). The genes were mapped to Gene Ontology (GO) terms for this purpose. The FDR cut-off associated with this analysis was set at <0.05 in order to identify significantly enriched functional terms within the networks.

Pathway enrichment analysis. The Web-based Gene Set Analysis Toolkit (WebGestalt) (29,30) has become one of the most popular software tools in pathway analysis. In this study, WebGestalt was used to perform pathway enrichment analysis of the genes comprised in the network of all DEGs and those comprised in the network of high degree-node DEGs. The genes were mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) database-derived pathways for this purpose. P<0.05 was set as the threshold for the identification of significantly enriched pathways.

Results

Differentially expressed genes. Raw data were normalized prior to differential expression analysis (Fig. 1A). The comparison between the control and disease group using overall data for all the samples is shown in Fig. 1B. The expression level of the disease group was significantly lower (P=3.05E-5) compared to the control group. The genes meeting the criteria for identification of DEGs (FDR <0.05, |logFC| >1) were selected. A total of 337 differentially expressed genes were identified, of which 185 were downregulated and 152 were upregulated in AD (data not shown).

Interaction networks of DEGs. Interaction networks were built with Osprey software in order to identify interactions among the DEGs. User-defined interactions were added or subtracted from mouse-over pop-up windows that link to the database (23). An initial network comprised a total of 206 DEGs (Fig. 2A). Networks of DEGs with different degrees of node were also built: Fig. 2B, C and D show the networks of DEGs with degree of node of ≥3, ≥5 and ≥10, respectively. The network that consisted of DEGs with degree of node ≥10 contained 41 genes, among which those coding for matrix
Table I. Functional enrichment analysis of Gene Ontology (GO) terms for (A) the network of all differentially expressed genes (DEGs) and (B) the network comprising DEGs with degree of node >10.

A, Network of all DEGs

<table>
<thead>
<tr>
<th>Biological process, GO term</th>
<th>No. of genes</th>
<th>P-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0008544: Epidermis development</td>
<td>15</td>
<td>2.95E-07</td>
<td>5.01E-04</td>
</tr>
<tr>
<td>0030216: Keratinocyte differentiation</td>
<td>9</td>
<td>3.60E-06</td>
<td>0.006099</td>
</tr>
<tr>
<td>0009913: Epidermal cell differentiation</td>
<td>9</td>
<td>6.98E-06</td>
<td>0.011834</td>
</tr>
<tr>
<td>0042493: Response to drug</td>
<td>14</td>
<td>1.07E-05</td>
<td>0.018058</td>
</tr>
<tr>
<td>0009991: Response to extracellular stimulus</td>
<td>14</td>
<td>1.30E-05</td>
<td>0.022048</td>
</tr>
<tr>
<td>0009725: Response to hormone stimulus</td>
<td>18</td>
<td>1.54E-05</td>
<td>0.026167</td>
</tr>
<tr>
<td>0009611: Response to wounding</td>
<td>22</td>
<td>1.64E-05</td>
<td>0.027821</td>
</tr>
<tr>
<td>0031667: Response to nutrient levels</td>
<td>13</td>
<td>2.08E-05</td>
<td>0.035279</td>
</tr>
<tr>
<td>0009615: Response to virus</td>
<td>10</td>
<td>2.15E-05</td>
<td>0.036422</td>
</tr>
<tr>
<td>0010033: Response to organic substance</td>
<td>26</td>
<td>2.41E-05</td>
<td>0.040820</td>
</tr>
<tr>
<td>0018149: Peptide cross-linking</td>
<td>6</td>
<td>2.62E-05</td>
<td>0.044354</td>
</tr>
<tr>
<td>0031424: Keratinization</td>
<td>7</td>
<td>2.72E-05</td>
<td>0.046187</td>
</tr>
</tbody>
</table>

B, Network of DEGs with degree of node >10

<table>
<thead>
<tr>
<th>Biological process, GO term</th>
<th>No. of genes</th>
<th>P-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0008544: Epidermis development</td>
<td>9</td>
<td>4.32E-08</td>
<td>6.69E-05</td>
</tr>
<tr>
<td>0031424: Keratinization</td>
<td>6</td>
<td>1.36E-07</td>
<td>2.10E-04</td>
</tr>
<tr>
<td>0018149: Peptide cross-linking</td>
<td>5</td>
<td>8.42E-07</td>
<td>0.001306</td>
</tr>
<tr>
<td>0030216: Keratinocyte differentiation</td>
<td>6</td>
<td>1.20E-06</td>
<td>0.001861</td>
</tr>
<tr>
<td>0009913: Epidermal cell differentiation</td>
<td>6</td>
<td>1.86E-06</td>
<td>0.002877</td>
</tr>
<tr>
<td>0030855: Epithelial cell differentiation</td>
<td>7</td>
<td>2.40E-06</td>
<td>0.003715</td>
</tr>
</tbody>
</table>

FDR, false discovery rate.

metalloproteinase 1 (MMP1), MMP9, small proline-rich repeat protein 1A (SPRR1A) and loricrin (LOR).

Functional enrichment analysis. Functional enrichment analysis using GO terms was performed for the network of all DEGs and the network of DEGs with degree of node >10, comprising 12 and 6 genes, respectively (Table IA and B). Epidermis development was the most significantly enriched functional term in the two networks (FDR =5.01E-04 in the network of high degree-node genes). This term was associated with a number of high node-degree DEGs, coding for LOR, type I cytokeratin 17 (KRT17), connective tissue growth factor (CTGF), SPRR2D, SPRR1A, SPRR1B, transglutaminase 1 (TGM1) and involucrin (IVL).

Pathway enrichment analysis. Genes from the network of all DEGs and the network of DEGs with degree of node >10 were subjected to pathway enrichment analysis using pathways derived from KEGG. According to the cut off of P<0.05, only one pathway, the chemokine signaling pathway, was enriched (Fig. 3) (P=0.0490978 for the work of all the DEGs and P=0.0399795 for the network of DEGs with degree of node >10). DEGs that belong to the chemokine pathway include chemokine (C-C motif) receptor 7 (CCR7), chemokine (C-C motif) ligand 19 (CCL19), signal transducer and activator of transcription 1 (STAT1) and phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1).

Discussion

Statistical comparison of gene expression profiles between AD and healthy skin samples was carried out in the present study to identify DEGs, followed by construction of DEGs networks, functional and pathway enrichment analyses, aiming to investigate the interactions and the functions of the identified DEGs. The long-term goal of the study was to identify AD-related key genes.

A range of genes related to epidermis development were identified among the high degree-node DEGs, such as genes coding for LOR, KRT17, KRT16, SPRR2D, SPRR1A, SPRR1B, and involucrin (IVL). Kim et al showed that the two proteins were downregulated in skin with AD compared to healthy skin (31). Those authors suggested that the two proteins were inhibited by tyrosine hydroxylase 2 (Th2) cytokines through STAT-6 (31), with a deficiency in LOR protein expression leading to a reduction in the expression of epidermal growth factor receptor (EGFR), e-cadherin, and occludin (32). In addition, Sevilla et al reported that mice deficient in IVL, envoplakin
Figure 1. (A) Boxplot of normalized gene expression data. The x-axis indicates the sample name and the y-axis represents the expression value. (B) Boxplot comparison between AD and healthy control group data, calculated over all the samples. ***P<0.001. Atopic dermatitis (AD) samples are in blue, while healthy control samples are in red.

Figure 2. Interaction networks for (A) all differentially expressed genes (DEGs), as well as DEGs with degree of node (B) >3, (C) >5 and (D) >10. N is the number of DEGs in each network.
and periplakin had a defective epidermal barrier (33). Thus, it can be hypothesized that LOR and IVL are important in maintaining the status of the skin/epidermal barrier.

KRT17 is a type I intermediate filament chain, expressed in nail beds, hair follicle, sebaceous glands and other epidermal appendages. In our study, KRT17 was identified as part of the network of DEGs with degree of node >10, and it was mapped to the GO biological process ‘epidermis development’. In a previous study, the expression of KRT17 in inflammatory skin diseases was shown to be regulated by lymphocytes via cytokine production (34). Other studies indicated that keratine intermediate filament proteins acted as regulators of inflammation and immunity in skin diseases (35), and that KRT17 promoted epithelial cell proliferation (36). Findings of a recent study suggested that the expression levels of KRT17 and IVL were increased in canine atopic dermatitis (37). Therefore, KRT17 may be a useful molecular target in AD treatment.

The genes coding for SPRRs locate near to the LOR and IVL genes, within a cluster of 1.5 Mbp on the chromosome 1q21, and most likely evolved from a common ancestor. Moreover, SPRR genes are strongly induced during differentiation of human epidermal keratinocytes (38). A study by Zimmermann et al indicated that SPRR2 encoded an allergen and was induced by IL-13 in experimental allergic responses that may be involved in disease pathophysiology (39). Notably, SPRR2A and SPRR2B proteins were considered to be involved in epithelial differentiation, but not allergic disease (39), prior to this study. In the present study, the genes coding for SPRR2D, SPRR1A and SPRR1B were found to be differentially expressed in AD and mapped to the enriched functional categories of epidermis development, keratinization, peptide cross-linking and keratinocyte differentiation. Thus, further studies are necessary to better understand the involvement of this family of proteins in AD pathogenesis.

Pathway enrichment analysis revealed that AD-related genes were significantly enriched for constituents of the chemokine signaling pathway, and the relevant proteins extensively interacted with other proteins, such as CCR7, CCL19, STAT1 and PIK3R1. The chemokine receptor CCR7 can control the migration of memory T cells to inflamed tissues, as well as stimulate dendritic cells maturation (40). CCL19 specifically binds to CCR7, while the involvement of the CC chemokine family

Figure 3. Chemokine signaling pathway, adapted from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways database. Differentially expressed genes identified in the present study are in the pink-color boxes.
proteins in AD has been reported in numerous studies (41-43). STAT1 is also involved in the chemokine signaling pathway: this protein is phosphorylated by receptor-associated kinases and translocates to the cell nucleus where it acts as a transcription activator. Heishi et al. reported that the STAT1 gene was differentially expressed in peripheral blood mononuclear cells from AD patients compared to controls and may thus be a useful marker for evaluating AD (44). The gene can be activated by IFNγ and is thus involved in immune-mediated inflammation. STAT1 was also upregulated in lesional compared to non-lesional AD (45). As for PIK3R1, encoding the α regulatory subunit of phosphoinositide 3-kinase (PI3K) (46), there are few reports concerning its role in AD, limited to mouse models of atopic disorders (47,48). In conclusion, the chemokine signaling pathway genes CCR7, CCL19 and STAT1 may be key genes in AD, and PIK3R1 may represent a new therapeutic target in AD.

In summary, we adopted a microarray approach to identify key genes associated with AD. A number of potential AD targets was identified, such as the DEGs coding for LOR, IVL, KRT17 and SPRRs, which are involved in epidermis development, as well as the DEGs encoding CCR7, CCL19 and STAT1, proteins of the chemokine signaling pathway. Among these DEGs, two represent potentially novel therapeutic targets for AD, PIK3R1 and KRT17. However, this study was limited by the small sample size, while we did not experimentally confirm that the identified DEGs are involved in AD. Future studies focusing on these genes may provide important contributions to the diagnosis and treatment of AD.

References


