Abstract. The aim of the current study was to investigate disease-associated genes and related molecular mechanisms of osteoarthritis (OA) and rheumatoid arthritis (RA). Using GSE7669 datasets downloaded from Gene Expression Omnibus databases, the differentially expressed genes (DEGs) between RA and OA synovial fibroblasts (SFBs) (n=6 each) were screened. DEG-associated co-expression and topological properties were analyzed to determine the rank of disease-associated genes. Specifically, the fold change of differentially expressed genes, the clustering coefficient and the degree of differential gene co-expression were integrated to determine the disease-associated gene ranking. The underlying molecular mechanisms of these crucial disease-associated genes were investigated by gene ontology (GO) enrichment analysis. A total of 1313 DEGs, including 1068 upregulated genes and 245 downregulated genes were observed. The top 20 disease-associated genes were identified, including proteoglycan 4, inhibin βB, carboxypeptidase M, alcohol dehydrogenase 1C and integrin β2. The major GO biological processes of these top 20 disease-associated genes were highly involved in the immune system, such as responses to stimuli, immune responses and inflammatory responses. This large-scale gene expression study observed disease-associated genes and their associated GO function in RA and OA, which may provide opportunities for biomarker development and novel insights into the molecular mechanisms of these two diseases.

Introduction

Human rheumatoid arthritis (RA), a polyarticular disease of autoimmune nature (1), is perpetuated by an invasive pannus tissue, whereas osteoarthritis (OA), as a non-inflammatory degenerative disease of the articular cartilage (2), is characterized by an increased tendency for novel blood vessel formation (3,4). Furthermore, recent studies have reported information regarding the differences in pathogenesis between RA and OA. Patients with RA present with joint destruction caused by hyperplasia of the synovial lining, infiltration of mononuclear cells into the sublining layer, stimulation of fibroblast-like synoviocytes and the increase of catabolic mediators, including interleukin (IL)-1β tumor necrosis factor (TNF)-α and matrix metalloproteinases (MMP) (5). By contrast, in patients with OA, joint destruction is due to cartilage degradation and elevated concentrations of cartilage matrix components, which elicit the presence of synovitis. In addition, synovitis aggravates the damage of articular cartilage by releasing inflammatory cytokines and destructive proteases (6).

Although OA and RA have different modes of pathogenesis, the current treatment of these two disease is similar, including nonsteroidal anti-inflammatory drugs, applied for pain and inflammation management (7,8); disease-modifying antirheumatic drugs, which function as a classical first-line therapy to minimize or prevent joint damage (7,9); and surgical treatment performed to replace the joints (10,11). However, these approaches induce a number of adverse events and less than satisfactory clinical outcomes (12,13). Thus, disease-specific therapy requires further investigation.

Genetic factors are also critical in the pathogenesis of RA and OA. Bramlage et al (14) identified that bone morphogenetic protein (BMP)-4 and BMP-5 were downregulated in OA and RA compared with that expressed in normal synovial tissue, suggesting a role of distinct BMPs in joint homeostasis, which may be altered in inflammatory and degenerative joint diseases. In addition, Pohlers et al (15) demonstrated that upregulation of the tumor growth factor (TGF)-β pathway was observed in RA synovial fibroblasts (SFBs), resulting in significant overexpression of MMP-11 mRNA and protein in RA SFBs, but not in OA SFBs. However, differentially expressed...
genes (DEGs) and molecular mechanisms underlying RA and OA are not yet fully understood.

In the current study, comparative analysis of DEG characteristics between RA and OA profiles was performed to identify DEGs with potential pathophysiological relevance. Differential gene co-expression networks were constructed and analyzed to identify disease candidate genes.

Materials and methods

Microarray data analysis. The gene expression data was downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) using GEO accession no. GSE7669 (15). The database contains six RA and six OA SFBs derived from a human study using the Affymetrix Human Genome U95 version 2 (Affymetrix Inc., Santa Clara, CA, USA) platform. Background-corrected signal intensities were determined using the MAS 5.0 software (Affymetrix®). The normalization of datasets obtained on Affymetrix arrays was performed using the preprocessCore package in R (16).

Screening DEGs. Significance analysis of microarray (SAM) is widely used to detect genes on a microarray with statistically significant changes in expression (17). As an alternative to a t-test, significance SAM4.0 was employed for determination of differentially expressed genes between RA and OA SFBs and the threshold was set as log |fold change| ≥2.

Constructing differential gene co-expression networks. Prior to the construction of co-expression networks, the expression value was determined from differentially expressed gene samples between RA and OA. Next, Pearson's coefficient was applied to calculate correlations between DEGs. The differential co-expressiond network was then constructed as the threshold was set as r≥0.8.

Network topological analysis. The Cytoscape plug-in Network Analyzer (18) was used for network visualization and gene co-expression analysis. For each network the number of nodes and edges was simply calculated. The clustering coefficient of a network $C_n$ was counted as the average clustering coefficient of all of its non-singleton nodes via the formula: $C_n = 2e_n/k_n(k_n-1)$ (1); where $e_n$ denotes the number of edges between the $k_n$ neighbors of $n$ and $k_n$ is the degree of $n$ (19).

Gene ranking. Conventionally, disease-associated gene ranking is determined by linkage analysis and gene expression profile analysis (20-22). However, these methods are mostly limited to a single statistic indicator. In the current study, the fold change of differentially expressed genes, the clustering coefficient and the degree of differential gene co-expression were integrated to determine the disease-associated gene ranking using the formula (2): $\text{Cri}_n = FC_n + \text{degree}_n + C_n$ (2); where $\text{Cri}_n$ is an indicator of the gene ranking (the larger the value is, the higher the gene ranks), $FC_n$ stands for the value of gene fold change, $\text{degree}_n$ and $C_n$ denotes the average degree and the clustering coefficient of a co-expression network respectively.

Calculating GO enrichment. Biological function and candidate gene-associated biological pathways can be determined by GO (http://www.geneontology.org). The open access software DAVID (23) was used to access the GO enrichment of candidate genes, which is a slightly modified Fisher's exact test, identical to the EASE score (24).

Results

Differentially gene expression. Gene expression in six RA SFBs was compared with that in six OA SFBs. The R preprocessCore was used to normalize and preprocess the presented data (log |fold changel ≥2) (Fig. 1). A total of 1313 differentially expressed genes were observed, in which 1068 genes were upregulated and 245 genes were downregulated in OA SFBs compared with RA SFBs (Fig. 2).
Differential co-expression network. To construct a differential gene co-expression network, data from a total 1313 DEGs were extracted from the expression profile and Pearson’s coefficient was applied to calculate correlations between these DEGs (the threshold as r≥0.8). A total 1302 nodes and 20372 edges were identified among the differential gene co-expression network, which is comprised of two closely connected sub-networks (Fig. 3).

Top 10 genes with the highest node degrees in the co-expression network. Clustering coefficient and degree were analyzed to detect the importance of disease-associated genes from differential gene co-expression networks. In the current study, the top 10 degree of corresponding genes are shown in Table I. Briefly, tubulin folding cofactor, ATP-binding cassette sub-family A member 3 and dimethylarginine dimethylaminohydrase 2 were the top three genes of the list, with a considerably higher degree in the co-expression network.

Top 20 disease candidate genes. To determine the association between DEGs and the two diseases, OA and RA, disease candidate genes were ranked according to the \( \text{Cri}_n \) value using the formula \( \text{Cri}_n = FC_n + \text{degree}_n + C_n \). As presented in Table II, the top 5 of 20 disease candidate genes were inhibin \( \beta \) B (INHBB), carboxypeptidase M, alcohol dehydrogenase 1C, integrin \( \beta_2 \) (ITGB2) and collagen, type XI, \( \alpha_1 \) (COL11A1), respectively.

Functional annotation of candidate genes. The analysis of GO enrichment was used to detect the association between the top 20 disease-associated candidate genes and associated biological processes and pathways (P<0.05). Overall, the majority of the top 20 disease-associated candidate genes, including INHBB, COL11A1, and close homolog of L1, were intensively enriched in immune-associated biological process terms, including defense, inflammatory, immune responses, immune system process and response to wounding. In addition, ITGB2, THBS1, COL11A1, and close homolog of L1 were closely involved in cell adhesion and biological adhesion (Table III).

Discussion

Gene expression studies have been widely used to allow improved diagnosis and identify novel pathways implicated in the pathogenesis of autoimmune diseases. In the current study, DEGs in OA SFBs compared with RA SFBs were identified based on gene expression profiling, 1068 upregulated genes and 206 downregulated genes, which are encompassed in the top 20 disease-associated candidate genes, including INHBB, COL11A1, and close homolog of L1, were intensively enriched in immune-associated biological process terms, including defense, inflammatory, immune responses, immune system process and response to wounding. In addition, ITGB2, THBS1, COL11A1, and close homolog of L1 were closely involved in cell adhesion and biological adhesion (Table III).
et al. : ANALYSIS OF DEGs BETWEEN RA AND OA

122 and 245 downregulated genes were observed. Similarly, a previous study has reported different biological properties between RA and OA SFBs. Higher levels of specific cytokines were found in RA SFBs compared with OA SFBs, including epidermal growth factor, basic fibroblast growth factor, TGF-β1, granulocyte-macrophage colony-stimulating factor, and THBS1.

Table III. Top 20 disease candidate genes and related BP terms.

<table>
<thead>
<tr>
<th>Rank</th>
<th>BP term</th>
<th>N</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Response to stimulus</td>
<td>9</td>
<td>INHBB, PRG4, CCR3, RSAD2, MS4A2, ITGB2, THBS1, COL11A1, ABCA3</td>
</tr>
<tr>
<td>2</td>
<td>Localization</td>
<td>8</td>
<td>SLC29A1, MYRIP, CLIC5, MS4A2, ITGB2, THBS1, ABCA3, DNM1</td>
</tr>
<tr>
<td>3</td>
<td>Defense response</td>
<td>6</td>
<td>INHBB, CCR3, RSAD2, MS4A2, ITGB2, THBS1</td>
</tr>
<tr>
<td>4</td>
<td>Response to stress</td>
<td>6</td>
<td>INHBB, CCR3, RSAD2, MS4A2, ITGB2, THBS120</td>
</tr>
<tr>
<td>5</td>
<td>Response to external stimulus</td>
<td>6</td>
<td>INHBB, CCR3, MS4A2, ITGB2, THBS1, COL11A1</td>
</tr>
<tr>
<td>6</td>
<td>Immune system process</td>
<td>5</td>
<td>PRG4, RSAD2, MS4A2, ITGB2, THBS1</td>
</tr>
<tr>
<td>7</td>
<td>Cell adhesion</td>
<td>5</td>
<td>CCR3, ITGB2, THBS1, COL11A1, CHL1</td>
</tr>
<tr>
<td>8</td>
<td>Biological adhesion</td>
<td>5</td>
<td>CCR3, ITGB2, THBS1, COL11A1, CHL1</td>
</tr>
<tr>
<td>9</td>
<td>Immune response</td>
<td>4</td>
<td>PRG4, RSAD2, MS4A2, ITGB2, THBS1</td>
</tr>
<tr>
<td>10</td>
<td>Inflammatory response</td>
<td>4</td>
<td>CCR3, MS4A2, ITGB2, THBS1</td>
</tr>
<tr>
<td>11</td>
<td>Response to wounding</td>
<td>4</td>
<td>CCR3, MS4A2, ITGB2, THBS1</td>
</tr>
<tr>
<td>12</td>
<td>Inner ear morphogenesis</td>
<td>2</td>
<td>CLIC5, COL11A1</td>
</tr>
</tbody>
</table>

Rank, the rank of BP term determined by the number of the 20 disease-related genes enriched in the BP term. N, number of the 20 disease-related genes involved in the BP term. BP, biological process; INHBB, inhibin β B; PRG4, proteoglycan 4; RSAD2, radical S-adenosyl methionine domain containing 2; MS4A2, membrane-spanning 4-domains, subfamily A, member 2; ITGB2, integrin β2; THBS1, thrombospondin 1; COL11A1, collagen, type XI, α 1; ABCA3, ATP-binding cassette sub-family A member 3; SLC29A1, solute carrier family 29; MYRIP, myosin-VIIa- and Rab-interacting protein; CLIC5, chloride intracellular channel 5; DNM1, dynamin 1; THBS120, thrombospondin 120.

Figure 3. Differential gene co-expression network. A node represents one DEG (RA vs. OA) and is interconnected in the network. Co-expressed genes are linked by edges. DEGs, differentially expressed genes; RA, rheumatoid arthritis; OA, osteoarthritis.
IL-1β and IL-6. In addition, by contrast with the OA SFBs, RA SFBs were observed to stimulate [3H]thymidine incorporation in the murine fibroblast cell line (25). Therefore, the current results support the hypothesis that RA and OA may result in the alterations of gene expression in the SFBs.

The development of an RA and OA gene co-expression network based on the topological analysis is critical, since it may provide visualized structural information regarding the connectivity of genes, compared with the traditional clustering analysis (26). Notably, the clustering coefficient and degree are two of the most important features of the network model. In the present study, the top 10 clustering coefficient Cri, associated genes were identified, which indicates a highly significant association between these genes and disease status. Notably, to the best of our knowledge the present study was the first to use the formula: Cri = FCn + degree + Cn, which is comprised of fold change, average degree and clustering coefficient of co-expression gene network to calculate the top 20 disease-associated candidate genes.

For OA disease, Martin et al (27) reported that the single nucleotide polymorphism, rs2615977, located in intron 31 of COL11A1 (5th gene at ranking) is highly associated with OA. In addition, COL11A1 has been used as a significant target for musculoskeletal disease research (28). In the present study, COL11A1 was observed to be downregulated in OA samples compared with RA samples, which further supports the results of a previous study that suggested COL11A1 may have be associated with OA (29). In addition, the inhibition of cartilage hyperplasia appears to be another favorable approach to relieve symptoms of OA disease. Ruan et al (30) demonstrated that PRG4 (9th gene at ranking) is highly associated with OA (29). Notably, ITGB2 was identified to be upregulated in OA rather than RA, which further supports the hypothesis that ITGB2 may have a potential association with the pathogenesis of RA and OA.

In RA, Rinaldi et al (31) suggested that β1 integrins contribute to the tight binding of RA SFBs to the matrix and regulate extracellular matrix remodeling in the RA disease process in vivo. The function of β1 integrins is similar to that of the ITGB2 identified in this study (4th gene at ranking). Notably, ITGB2 was identified to be upregulated in RA rather than OA, which further supports the hypothesis that ITGB2 may be critical in facilitating the RA disease process. INHBA and INHBB (1st gene at ranking) are two subunits of inhibin. El-Gendi et al (32) demonstrated that the serum level of inhibin β A was significantly higher in 60 patients with RA compared with 20 normal patients. In the present study, the expression level of INHBB was identified to be higher in RA compared with OA, indicating that INHBB may have a similar function to INHBA and may serve as a novel biological marker in the diagnosis of RA disease.

Furthermore, functional enrichment analysis of the top 20 disease-associated genes was performed to demonstrate the possible biological mechanisms underlying AR and OA, including response to stimulus, localization, response to stress, response to external stimulus and process involved in the development of the immune system. This finding is consistent with the results of previous studies, suggesting that the immune system is pivotal in auto-inflammatory and non-auto-inflammatory arthritis. The immune system regulates the alteration of cell osmosis and the inflammatory response caused by periostitis (33-36). The majority of the other top 20 disease-associated genes were enriched in terms of cell and biological adhesion. Notably, cell adhesion is also highly correlated with the pathogenesis of AR and OA. Karatay et al (37) stated that the decreased intracellular adhesion molecule-1 and vascular cell adhesion molecule-1 levels following intra-articular hyaluronic acid (HA) injection may aid in explaining the anti-inflammatory effects of HA therapy in OA of the knee.

In conclusion, the present study offers significant information that may aid in understanding the molecular mechanisms underlying OA and RA. The top 20 disease-associated genes and associated BP terms were observed in this study, which may facilitate the design of targeted therapy for OA and RA in the near future.

Acknowledgments

This study was supported by the Shanghai Medical Key Subject Construction Project (ZK2012A28) and National Clinical Key Specialty Construction Project.

References


