Regulation different network analysis of rheumatoid arthritis (RA) and osteoarthritis (OA)


Department of Orthopedics, Shanghai East Hospital, Tongji University, Pudong New district, Shanghai, China

Feng Yin, Shu-lin Luo and Feng Yuan should be regarded as co-first authors

Abstract. – BACKGROUND: Rheumatoid arthritis (RA) is a systemic autoimmune disease characterized by chronic synovitis that progresses to destruction of cartilage and bone.

AIM: The purpose of this study was to employ microarray analysis combined with bioinformatics techniques to evaluate differential gene expression in BM-derived mononuclear cells obtained from patients with rheumatoid arthritis (RA) or osteoarthritis (OA) to study the pathogenesis of this disease.

MATERIALS AND METHODS: Gene expression profiles in BM-derived mononuclear cells from 9 RA and 10 OA patients were obtained from GEO.

RESULTS: The bone marrow (BM) mononuclear cells showed 2581 up-regulated and 649 down-regulated genes in RA patients relative to the OA group: Our analysis indicated that several differentially expressed genes might play crucial roles in RA development, including SP1, RARA, ETS1, ETS2, FOS and ESR1.

CONCLUSIONS: Further analysis predicted these genes might be involved in RA through cancer related pathways and immunity related pathways. Furthermore, these genes may serve as novel therapeutic targets for the treatment of RA.

Key Words: Regulation network, Rheumatoid arthritis, Osteoarthritis.

Introduction

Rheumatoid arthritis (RA) is a chronic, systemic inflammatory disorder that may affect many tissues and organs, but principally attacks flexible (synovial) joints1, resulting in a painful condition and even disability if inadequately treated. RA is one of the most common systemic autoimmune diseases affecting an estimated 20 million individuals worldwide. The annual global prevalence rate of RA is approximately 1-2%2, with incidence of approximately 3 per 100,000 new cases per year. Much is known about the causes of RA. It is up to three times more common in smokers than non-smokers, particularly in men, heavy smokers, and those who are rheumatoid factor positive3.

Substantial progress has been made in the pathogenesis of RA. Genetic studies have identified genes with important functions related to transcription factors, cytokines, costimulatory molecules, intracellular signaling pathways, etc. during the past decades. In a series of studies, nuclear factor κB (NF-κB) and activator protein-1 (AP-1) were found to be abundantly expressed in RA synovial tissue, and their levels were significantly higher in RA than in osteoarthritis (OA) or normal synovial tissues4,5. A lot of genomic regions, such as TNFAIP36,7, TRAF1-C58, CD409, STAT410, CTLA411, etc. have been demonstrated strongly associated with RA in the genome-wide association studies. These discoveries highlight the importance of the CD40/NF-κB signaling pathway, Janus associated kinase/signal transducer and activator of transcription (Jak/STAT) signaling pathway and other pathways in RA. However, the relative influence of each of the pro- and anti-inflammatory mediators is likely to be complex and requires considerable further analysis.

The advent of DNA microarray technology, which facilitates survey to identify genes and biological pathways that are associated with clinically defined conditions12, has provided a powerful tool to understand the complex pathogenesis and heterogeneous manifestations of autoimmune diseases including RA. DNA microarray has been used to investigate the pathogenesis of juvenile idiopathic arthritis (JIA) and systemic lupus erythematosus (SLE) by previous studies13,14. It is demonstrated that DNA microarray is more effective when combined with bioinformatics techniques such as gene ontology (GO) databases and pathway analysis software.

In this study, we identified the differentially expressed genes based on their gene expression
profiles in bone marrow (BM)-derived mononuclear cells from RA patients comparing with those of osteoarthritis (OA) patients, and constructed a regulatory network by cytoscape. We also performed GO and pathway enrichment analysis to describe the function of genes in our regulatory network.

**Materials and Methods**

**Data Source**

*Affymetrix Microarray Data*

The transcription profile of GSE27390 was obtained from the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) database (http://www.ncbi.nlm.nih.gov/geo/) which is based on the Affymetrix Human Genome U133 Plus 2.0 Array. Total 19 chips, purchased from Laboratory of Immune Regulation (567-0085, Japan), were used for analysis. Gene expression profiles of bone marrow mononuclear cells (BMMC) from patients with rheumatoid arthritis (RA) (n=9) was compared with those from patients with osteoarthritis (OA) (n=10). The bone marrow samples from the OA patients are used as controls. Up- and down-regulated genes were identified by comparing the Gene expression profiles (GEPs) from the two patient groups.

**Regulation Data**

774 pairs of regulatory relationship between 219 transcription factors (TFs) and 265 target gene were collected from TRANSFAC database (http://www.gene-regulation.com/pub/databases.html) 5722 pairs of regulatory relationship between 102 transcription factors (TFs) and 2920 target genes were collected from TRED (transcriptional regulatory element database) (http://rulai.cshl.edu/TRED/).

Combined the two regulation datasets, total 6328 regulatory relationships between 276 TFs and 3002 target genes were collected.

**Methods**

**Differentially Expressed Genes Analysis**

For the GSE27390 dataset, the limma method was used to identify Differentially expressed genes (DEGs). The original expression datasets from all conditions were extracted into expression estimates, and then constructed the linear model. The DEGs only with the fold change value larger than 2 and p-value less than 0.05 were selected.

**Co-expression Analysis**

For demonstrating the potential regulatory relationship, the Pearson Correlation Coefficient (PCC) was calculated for all pair-wise comparisons of gene-expression values between TFs and the DEGs. The regulatory relationships whose absolute PCC are larger than 0.5 were considered as significant.

**Regulation Network Construction**

Using the regulation data that have been collected from TRANSFAC database and TRED database, we matched the relationships between differentially expressed TFs and its differentially expressed target genes.

Base on the above two regulation datasets, the significant relationships (PCC > 0.5 or PCC < –0.5) between TFs and its target genes, we constructed the regulation networks by Cytoscape.

**Gene Ontology Analysis**

DAVID (Database for Annotation, Visualization and Integrated Discovery), a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. We used the DAVID to identify over-represented Gene Ontology (GO) categories in biological process.

**Significance Analysis of Pathway**

We adopted an impact analysis that includes the statistical significance of the set of pathway genes but also considers other crucial factors such as the magnitude of each gene’s expression change, the topology of the signaling pathway, their interactions, etc. In this model, the Impact Factor (IF) of a pathway Pi is calculated as the sum of two terms:

$$IF(Pi) = \log \left( \frac{1}{pi} \right) + \sum_{g \in Pi} \frac{|PF(g)|}{|\Delta E| \cdot N_{deg}(Pi)}$$

The first term is a probabilistic term that captures the significance of the given pathway Pi from the perspective of the set of genes contained in it. It is obtained by using the hyper geometric model in which $p_i$ is the probability of obtaining
at least the observed number of differentially expressed gene, $N_{de}$, just by chance.\(^{(23,24)}\)

The second term is a functional term that depends on the identity of the specific genes that are differentially expressed as well as on the interactions described by the pathway (i.e., its topology).

The second term sums up the absolute values of the perturbation factors (PFs) for all genes on the given pathway $P$.

The PF of a gene $g$ is calculated as follows:

$$PF(g) = \Delta E(g) + \sum_{u \in US_g} \frac{PF(u)}{N_{ds}(u)}$$

In this equation, the first term $E(g)$ captures the quantitative information measured in the gene expression experiment. The factor $E(g)$ represents the normalized measured expression change of the gene $g$. The first term $E(g)$ in the above equation is a sum of all PFs of the genes $u$ directly upstream of the target gene $g$, normalized by the number of downstream genes of each such gene $N_{ds}(u)$, and weighted by a factor $\beta_{ug}$, which reflects the type of interaction: $\beta_{ug} = 1$ for induction, $\beta_{ug} = -1$ for repression (KEGG supply this information about the type of interaction of two genes in the description of the pathway topology). $US_g$ is the set of all such genes upstream of $g$. We need to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, $N_{de}(P)$. In order to make the IFs as independent as possible from the technology, and also comparable between problems, we also divide the second term in equation 1 by the mean absolute fold change $\Delta E$, calculated across all differentially expressed genes.

KEGG (Kyoto Encyclopedia of genes and genomes) pathway analysis based on a system biology approach (http://vortex.cs.wayne.edu/) was used for pathway level analysis with the $p$-value less than 0.05.

**Results**

**Microarray Data Analysis in RA and OA Patients**

Publicly available microarray data set GSE27390 was obtained from GEO. Total 3230 genes with the fold change value larger than 2 and $p$-value less than 0.05 were selected as differentially expressed genes in BMMC from patients with RA compared with those from patients with OA. Among these genes, 2581 genes were up-regulated and the remaining 649 genes were down-regulated.

**Regulation Network Construction**

To obtain regulation network, based on the significant relationships (PCC > 0.5 or PCC < −0.5) between TFs and its target genes, 98 expression relationships including 40 TFs and 79 target genes were selected. By integrating expression relationships above, a regulation network of OA and RA was built between its TFs and target genes (Figure 1). In this network, the TFs SP1, RARA, ETS1, ETS2, FOS, E2F2 and ESR1 with higher degree form a local network which suggesting that these TFs may play an important role in the development of RA. FOS was down-regulated by BAG1, ESR2 and CEBPB while up-regulated by ESR1. Besides, TP53, PPARG, ESR1 down-regulated downstream target genes EGF R was also observed in our network.

**Function Analysis of the Network**

To describe the function of the regulation network, we performed GO enrichment analysis and pathway analysis. Several Gene Ontology (GO) categories were enriched among these genes in the regulation network, including regulation of transcription from RNA polymerase II promoter (GO: 0006357), positive regulation of gene expression (GO: 0010628) and positive regulation of cellular biosynthetic process (GO: 0031328) and so on (Table I). We also find that most of genes in our regulation network are positive regulated. Table I only lists the top 15 enriched GO terms.

To gain further insights into the function of genes in our regulation network, we performed pathway enrichment analysis using the KEGG pathways. Several KEGG pathways were enriched among these pathways in the regulation network, including p53 signaling pathway, graft-versus-host disease and pathways in cancer and so on (Table II). Table II only lists top 10 enriched KEGG pathways.

**Discussion**

Rheumatoid arthritis (RA) is a frequent chronic inflammatory disease characterized by distal,
bilateral and symmetrical lesions. The genetic contribution to RA has been recognized for decades, based on the results of twin and family studies\textsuperscript{25,26}. However, our understanding of the underlying mechanisms and pathogenesis of RA are still incomplete until recently. Therefore, it is important to identify the genes that contribute to RA risk to elucidate disease mechanisms and pathogenesis, thus informing the development of better diagnostic and therapeutic tools for this chronic and disabling disease. In this study, we constructed a regulatory network in BM-derived mononuclear cells (BMMC) from RA patients by analyzing differentially expressed genes based on their gene expression profiles with those of osteoarthritis (OA) patients. We aimed to employ microarray analysis combined with bioinformatics techniques such as gene ontology databases.

**Figure 1.** Regulation network of different expressed genes between OA and RA. The rhombic nodes stand for TFs and the circular nodes stand for target genes. The black lines stand for up-regulation, and the gray lines stand for down-regulation.

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Count</th>
<th>p-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0006357</td>
<td>Regulation of transcription from RNA polymerase II promoter</td>
<td>38</td>
<td>1.34E-19</td>
<td>2.30E-16</td>
</tr>
<tr>
<td>GO:0010628</td>
<td>Positive regulation of gene expression</td>
<td>32</td>
<td>6.17E-17</td>
<td>1.89E-13</td>
</tr>
<tr>
<td>GO:0031328</td>
<td>Positive regulation of cellular biosynthetic process</td>
<td>34</td>
<td>1.01E-16</td>
<td>1.89E-13</td>
</tr>
<tr>
<td>GO:0051173</td>
<td>Positive regulation of nitrogen compound metabolic process</td>
<td>33</td>
<td>1.36E-16</td>
<td>1.89E-13</td>
</tr>
<tr>
<td>GO:0009891</td>
<td>Positive regulation of biosynthetic process</td>
<td>34</td>
<td>1.55E-16</td>
<td>1.89E-13</td>
</tr>
<tr>
<td>GO:0045893</td>
<td>Positive regulation of transcription, DNA-dependent</td>
<td>29</td>
<td>2.30E-16</td>
<td>3.77E-13</td>
</tr>
<tr>
<td>GO:0045941</td>
<td>Positive regulation of transcription</td>
<td>31</td>
<td>2.36E-16</td>
<td>3.77E-13</td>
</tr>
<tr>
<td>GO:0051254</td>
<td>Positive regulation of RNA metabolic process</td>
<td>29</td>
<td>2.84E-16</td>
<td>5.77E-13</td>
</tr>
<tr>
<td>GO:0051252</td>
<td>Regulation of RNA metabolic process</td>
<td>52</td>
<td>4.17E-16</td>
<td>7.66E-13</td>
</tr>
<tr>
<td>GO:0006355</td>
<td>Regulation of transcription, DNA-dependent</td>
<td>51</td>
<td>7.51E-16</td>
<td>1.33E-12</td>
</tr>
<tr>
<td>GO:0010604</td>
<td>Positive regulation of macromolecule metabolic process</td>
<td>36</td>
<td>1.74E-15</td>
<td>3.05E-12</td>
</tr>
<tr>
<td>GO:0010557</td>
<td>Positive regulation of macromolecule biosynthetic process</td>
<td>32</td>
<td>1.87E-15</td>
<td>3.24E-12</td>
</tr>
<tr>
<td>GO:0045935</td>
<td>Positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process</td>
<td>31</td>
<td>4.11E-15</td>
<td>7.06E-12</td>
</tr>
<tr>
<td>GO:0045944</td>
<td>Positive regulation of transcription from RNA polymerase II promoter</td>
<td>25</td>
<td>5.62E-15</td>
<td>9.73E-12</td>
</tr>
<tr>
<td>GO:0042127</td>
<td>Regulation of cell proliferation</td>
<td>34</td>
<td>6.68E-15</td>
<td>1.14E-11</td>
</tr>
</tbody>
</table>
and pathway analysis software to evaluate differential gene expression in BMMC to study the pathogenesis of this disease.

From the result of regulatory network construction, we can find that the TFs SP1, RARA, ETS1, ETS2, FOS and ESR1 with higher degree form a local network which suggesting that these TFs may play an important role in the development of RA. Some of them were proved to be related with RA by previous study.

Sp1 (specificity protein 1) is ubiquitously expressed members of the Sp family of transcription factors that are involved in the expression and regulation of many genes, including tissue-specific expressed genes, viral genes, housekeeping genes, and cell cycle-regulated genes\textsuperscript{27,28}. It is known that Sp1 bind to the GC (guanine-cytosine content)-rich region in the promoter of a number of genes and regulate their promoter activities. Previous study\textsuperscript{29} indicated that Sp1 binding to the GC-rich region within its promoter region of the IL-1\textsubscript{a} gene was critical in RA. IL-1\textsubscript{a} is reportedly predominant in the early phase of arthritis and participates in the inhibition of proteoglycan synthesis\textsuperscript{30}. These findings suggested that Sp1 play an important role by binding to the GC-rich region in RA.

Retinoic acid receptors (RAR) are expressed in inflammatory cells and play an important role in cell proliferation and differentiation, as well as in regulation of cytokine and matrix metalloproteinase (MMP) production through ligand binding\textsuperscript{31,32}. RARA (retinoic acid receptor, alpha) inhibits the gene expression of collagenase by fibroblasts and monocytes\textsuperscript{32,33} transforming growth factor (TGF)-\textbeta\textsuperscript{34} and interleukin (IL)-6. Previous study\textsuperscript{35} observed the presence of RAR in synovial membrane (SM) of patients with OA and RA and suggested that RARs may play a role in the immunomodulation of synovial inflammation. Identically, our findings suggested that RARA play an important role in BMMC from RA patients. Therefore, RARA can be a potential target of therapeutic intervention in these arthritides.

ETS1 and ETS2 are members of ETS family of transcription factors. The encoded proteins function either as transcriptional activators or repressors of numerous genes, and are involved in stem cell development, cell senescence and death, and tumorigenesis. The ETS 1 transcription factor has been linked to the regulation of angiogenesis under both physiological and pathological conditions. Nicolas et al\textsuperscript{36} suggested ETS 1 transcription factor play important roles in the regulation of inflammatory angiogenesis in RA\textsuperscript{36}. Redlich et al\textsuperscript{37} suggested that ETS1 may be an important factor in the cytokine-mediated inflammatory and destructive cascade characteristic of RA. They observed that ETS-1 over-expressed in RA synovial tissue and this appears to be caused by TNF-alpha and IL-1\textsuperscript{37}. ETS-2 over-expression was also observed by previous study which suggesting the important role in the development and progression of RA\textsuperscript{38}.

Statistics demonstrated that RA develops two to four times more likely in women than men\textsuperscript{39,40}, therefore, sex hormones including androgens, estrogen, and progesterone may be related to this disparity\textsuperscript{41,42}. ESRs (Estrogen receptors) were investigated frequently to study the mechanistic basis of the gender difference. There is compelling evidence that estrogens and their receptors (ESRs) can play a role in the high prevalence of RA in females\textsuperscript{43}. However, there were also some study do not support the opinion that ESR2 genes are important to RA risk in women\textsuperscript{44}. In our result, ESR1 acted as a hub node that down-regulated epidermal growth factor receptor (EGFR), CYR61 (cysteine-rich61)
and RARA while up-regulated FOS (fructo-oligosaccharides), TGFA (transforming growth factor-alpha) and PNRC1 (protein-rich nuclear receptor coactivator 1). We also observed that ESR1 and ESR2 both act on FOS. Our result confirmed that ESRs may play an important role in the development of RA.

FOS encoded a protein c-Fos. c-Fos is a component of AP-1 (transcription factor activator protein-1). AP-1 regulates many genes that participate in RA, including TNF-α and metalloproteinas. High levels of AP-1 DNA binding activity are detected in nuclear extracts of RA synovial tissue compared with osteoarthritis. AP-1 decoy oligonucleotides suppress collagen-induced arthritis and inhibit IL-1, IL-6, and TNF-α. Matrix metalloproteinas-3 and MMP-9 production by synovial tissue. c-Fos are highly expressed in RA synovium, especially in the nuclei of cells in the intimal lining layer.

Besides, our results showed some of these genes might play important roles in RA through cancer related pathways, including pathways in cancer, p53 signaling pathway, and small cell lung cancer, etc. Pathways related to immunity are another significant category in our result, such as graft-versus-host disease, leukocyte transendothelial migration, antigen processing and presentation and natural killer cell mediated cytotoxicity. Our result suggested that apoptosis autoimmunity plays a pivotal role in the pathology of RA.

Conclusions

We employed microarray analysis combined with bioinformatics techniques to evaluate differential gene expression in BM-derived mononuclear cells from patients with RA and patients with OA to study the pathogenesis of this disease. Total 3230 genes were identified expressed differentially in our method and part of them were used to construct a regulatory network. Our analysis indicated several differentially expressed genes might play crucial roles in RA development, including SP1, RARA, ETS1, ETS2, FOS and ESR1. Further, our results predicted these genes might be involved in RA through cancer related pathways and immunity related pathways. Further research surrounding the role of genes in our result in RA is certainly needed, considering the lack of effective therapies currently available.

Conflict of Interest

The Authors declare that they have no conflict of interests.

References


30) van den Berg WB. Joint inflammation and cartilage degradation may occur uncoupled. Springer Semin Immunopathol 1998; 20: 149-164.


