Transcriptome network analysis reveals potential candidate genes for ankylosing spondylitis

Z.-Q. ZHU1,2, J.-S. TANG1, X.-J. CAO1

1Department of Orthopaedics, the First Affiliated Hospital of Nanjing Medical University, Nanjing, JiangSu Province, China
2Department of Orthopaedics, the Second Affiliated Hospital of Xuzhou Medical College, Xuzhou, JiangSu Province, China

Abstract. – OBJECTIVES: Ankylosing spondylitis (AS) is a chronic, inflammatory arthritis and autoimmune disease. 

BACKGROUND: The main symptom of AS is inflammatory spinal pain; with time, some patients develop ankylosis and spinal immobility. We aim to find cure available for ankylosing spondylitis. 

MATERIALS AND METHODS: We used the GSE11886 series to identify potential genes that related to AS to construct a regulation network. 

RESULTS: In the network, some of TFs and target genes have been proved related with AS in previous study, such as NFkB1, STAT1, STAT4, TNFSF10, IL3RA, and IL3RB. We also found some new TFs (Transcription Factors) and target genes response to AS, such as BXDC5, and EGFR. Further analysis indicated some significant pathways are associated with AS, including antigen processing and presentation and cytokine-cytokine receptor interaction, etc.; although not significant, there was evident that they play an important role in AS progression, such as apoptosis and systemic lupus erythematosus. 

CONCLUSIONS: Therefore, it is demonstrated that transcriptome network analysis is useful in identification of the candidate genes in AS.

Key Words: 
Ankylosing spondylitis, Transcription regulation network, Pathway regulation network, Microarray.

Introduction

Ankylosing spondylitis (AS) is a chronic, inflammatory arthritis and autoimmunity disease. The main symptom of AS is inflammatory spinal pain; with time, some patients develop ankylosis and spinal immobility. Damages to other organs also involved such as anterior uveitis, psoriasis and chronic inflammatory bowel disease. Ankylosing spondylitis is generally observed in young people, and is slightly prevail in men than women. In addition, it shows worse damage and more severe radiographic changes to male patients with longstanding AS.

There is no definite cure available for ankylosing spondylitis till now. Nonsteroidal anti-inflammatory drugs, sulfasalazine and immunomodulating agents have been used to reduce the pain and deterioration.

Ankylosing spondylitis is highly heritable arthropathy, the pathogenesis of AS is poorly understood. The association of HLA-B27 with ankylosing spondylitis (AS) has been known for over 33 years which approximately 90% of AS patients express the HLA-B27 genotype, and it remains one of the best examples of a disease association with a hereditary marker. HLA-B27 has been demonstrated to commonly form homodimers. It is postulated that these homodimers may lead to AS by abnormal peptide presentation or by abnormal recognition by NK cells. Genes IL1, ERAP1 and IL23 associated with the AS disease was also been demonstrated.

DNA microarray analysis as a global approach is applied to investigate physiological mechanisms in health and disease. Microarray experiments also have been designed to analyze genetic expression patterns and identify potential target genes of AS. These identified differentially expressed genes may provide insights into the pathogenesis of ankylosing spondylitis.

Therefore, the purpose of this paper is to propose the hypothesis that a transcriptome network can be developed such that a set of transcription factors, regulated the differently expression genes are induced by AS can be identified and modulated to it. Further analysis the genes and pathways in the network to identify potential mechanisms which response to the AS. The study does not address regulation network but searches for the significance pathways related to AS.

Materials and Methods

Affymetrix Microarray Data

One transcription profiles of Ischemic cardiomyopathy GSE11886 were obtained from a
public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/). Only 17 chips are usable for analysis. Samples were derived from the peripheral blood of 8 AS patients (median disease duration 13 yrs, range 1-43 yrs) and 9 healthy controls over 7 days with the use of granulocyte-macrophage colony-stimulating factor (GM-CSF). RNA quality was verified using an Agilent Bioanalyzer 2100 (Auburn, CA, USA), then reverse transcribed, converted to biotinylated cRNA using standard Affymetrix protocols, and hybridized to Affymetrix microarrays (HG-U133 Plus 2.0) according to protocols used by the Cincinnati Children’s Research Foundation Affymetrix GeneChip core. The HG-U133 Plus 2.0 microarray uses 54,120 probe sets to represent 38,572 UniGenes.

**Pathway Data**
KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals. The PATHWAY database records networks of molecular interactions in the cells, and variants of them specific to particular organisms (http://www.genome.jp/kegg/). Total 130 pathways, involving 2287 genes, were collected from KEGG.

**Regulation Data**
There are approximately 2600 proteins in the human genome that contain DNA-binding domains, and most of these are presumed to function as transcription factors. These transcription factors are grouped into 5 super class families, based on the presence of conserved DNA-binding domains. TRANSFAC database contains data on transcription factors, their experimentally-proven binding sites, and regulated genes.

Transcriptional Regulatory Element Database (TRED) has been built in response to increasing needs of an integrated repository for both cis- and trans- regulatory elements in mammals. TRED has done the curation for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. The curation is currently focusing on target genes of 36 cancer-related TF families. 774 pairs of regulatory relationship between 219 transcription factors (TFs) and 265 target genes were collected from TRANSFAC (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 (http://rulai.cshl.edu/TRED/).

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

**Differentially Expressed Genes (DEGs) Selection**
For the GSE11886 dataset, the limma method was used to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the RMA (Robust Multi-array Average) method with the default settings implemented in Bioconductor, and then construct the linear model. The DEGs only with the fold change value larger than 1.5 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 0.6 were considered as significant.

**Gene Ontology Analysis**
The Biological Networks Gene Ontology tool (BiNGO) is an open-source Java tool to determine which Gene Ontology (GO) terms are significantly overrepresented in a set of genes. The BiNGO analysis was used to identify over-represented GO categories in biological process with the p-value < 0.01.

**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 and the pathway relationships of the target genes, we construct the regulation networks by Cytoscape. Base on the significant relationships (PCC > 0.6 or PCC < -0.6) between TFs and its target genes, 33 putative regulatory relationships were predicted between 7 TFs and 22 target genes.

**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)
DEGs to pathways to get the regulation network between TFs and pathways.

**Results**

**Regulation Network Construction in Ankylosing Spondylitis**

To get DEGs of ankylosing spondylitis, we obtained publicly available microarray data sets GSE11886 from GEO. After microarray analysis, 178 DEGs with the fold change > 1.5 of GSE11886 and p-value < 0.05 were selected. To get the regulatory relationships, the co-expressed value (PCC ≥ 0.6) was chose as the threshold. Finally, we got 14 regulatory relationships between 8 TFs and their 8 differently expressed target genes. By integrating the regulatory relationships above, a regulation network of ankylosing spondylitis was built between TFs and its target genes (Figure 1). In this network, STAT1 and NFκB1 with higher degrees form a local network which suggesting that these two TFs may play an important role in ankylosing spondylitis. STAT1 and NFκB1 both regulate the IRF1 and IFNG target genes directly. Besides, the EGFR was regulated by 4 TFs were observed in our network.

**GO Analysis of the Regulation Network in Ankylosing Spondylitis**

Several Gene Ontology (GO) categories were enriched among these genes in the regulatory network, including regulation of apoptosis, regulation of programmed cell death and regulation of cell death and so on (Table I) using the BiNGO with the corr p-value < 0.01.

** Significant Pathway in Ankylosing Spondylitis**

To identify the relevant pathways changed in AS, we used a statistical approach on pathway level. Significance analysis at single gene level may suffer from the limited number of samples and experimental noise that can severely limit the power of the chosen statistical test. Pathway can provide an alternative way to relax the significance threshold applied to single genes and may lead to a better biological interpretation. So, we adopted a pathway based impact analysis method that contained many factor including the statistical significance of the set of differentially expressed genes in the pathway, the magnitude of each gene’s expression change, the topology of the signaling pathway, their interactions and so on. The impact analysis method yields many sig-
Analysis of candidate genes

From the regulation network of AS, we could find that many TFs and target genes closely related with AS have been linked by our method. The STAT1 and NFKB1 were hub nodes in our transcriptome network (Figure 1). Some of TFs and target genes have been proved related to AS in previous study, such as STAT1, STAT4, NFKB1, TNFSF10, IL2RA, and IL2RB. Importantly, we also predicted some new TFs and target genes response to AS, such as BXDC5, and EGFRT. Further analysis indicated some significant pathway associated with AS, including antigen processing and presentation, cytokine-cytokine receptor interaction, systemic lupus erythematosus, and apoptosis. We would in detail discuss our findings as following.

NFKB1 is an important transcription regulator that is activated by various intra- and extra-cellular stimuli such as cytokines, ultraviolet irradiation, and viral products. Activated NFKB1 translocates into the nucleus and stimulates the expression of genes involved in a wide variety of biological functions including inflammation and immunity, deve-

**Table I. GO biological process analysis.**

<table>
<thead>
<tr>
<th>GO-ID</th>
<th>Description</th>
<th>Count</th>
<th>p-value</th>
<th>Corr p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>42981</td>
<td>Regulation of apoptosis</td>
<td>10</td>
<td>1.22E-09</td>
<td>4.47E-07</td>
</tr>
<tr>
<td>43067</td>
<td>Regulation of programmed cell death</td>
<td>10</td>
<td>1.33E-09</td>
<td>4.47E-07</td>
</tr>
<tr>
<td>10941</td>
<td>Regulation of cell death</td>
<td>10</td>
<td>1.44E-09</td>
<td>4.47E-07</td>
</tr>
<tr>
<td>7259</td>
<td>JAK-STAT cascade</td>
<td>4</td>
<td>4.02E-08</td>
<td>9.34E-06</td>
</tr>
<tr>
<td>48518</td>
<td>Positive regulation of biological process</td>
<td>12</td>
<td>5.13E-08</td>
<td>9.52E-06</td>
</tr>
<tr>
<td>48660</td>
<td>Regulation of smooth muscle cell proliferation</td>
<td>4</td>
<td>1.04E-07</td>
<td>1.60E-05</td>
</tr>
<tr>
<td>45997</td>
<td>Positive regulation of cell differentiation</td>
<td>6</td>
<td>1.32E-07</td>
<td>1.76E-05</td>
</tr>
<tr>
<td>21993</td>
<td>Positive regulation of metabolic process</td>
<td>9</td>
<td>1.53E-07</td>
<td>1.78E-05</td>
</tr>
<tr>
<td>48522</td>
<td>Positive regulation of cellular process</td>
<td>11</td>
<td>3.15E-07</td>
<td>3.25E-05</td>
</tr>
<tr>
<td>19221</td>
<td>Cytokine-mediated signaling pathway</td>
<td>4</td>
<td>6.89E-07</td>
<td>5.21E-05</td>
</tr>
</tbody>
</table>

**Table II. Pathway significant analysis.**

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Pathway name</th>
<th>Impact factor</th>
<th>% Pathway genes in input</th>
<th>Corrected gamma p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG</td>
<td>Antigen processing and presentation</td>
<td>112.763</td>
<td>2.247</td>
<td>1.21E-47</td>
</tr>
<tr>
<td>KEGG</td>
<td>Cytokine-cytokine receptor interaction</td>
<td>27.766</td>
<td>6.464</td>
<td>2.51E-11</td>
</tr>
<tr>
<td>KEGG</td>
<td>Cell adhesion molecules (CAMs)</td>
<td>13.655</td>
<td>5.97</td>
<td>1.72E-05</td>
</tr>
<tr>
<td>KEGG</td>
<td>T cell receptor signaling pathway</td>
<td>12.158</td>
<td>6.481</td>
<td>6.90E-05</td>
</tr>
<tr>
<td>KEGG</td>
<td>Allograft rejection</td>
<td>10.054</td>
<td>10.526</td>
<td>4.75E-04</td>
</tr>
<tr>
<td>KEGG</td>
<td>Graft-versus-host disease</td>
<td>9.934</td>
<td>9.524</td>
<td>5.30E-04</td>
</tr>
<tr>
<td>KEGG</td>
<td>VEGF signaling pathway</td>
<td>9.613</td>
<td>1.351</td>
<td>7.10E-04</td>
</tr>
<tr>
<td>KEGG</td>
<td>Type 1 diabetes mellitus</td>
<td>9.491</td>
<td>9.091</td>
<td>7.92E-04</td>
</tr>
<tr>
<td>KEGG</td>
<td>Primary immunodeficiency</td>
<td>9.448</td>
<td>11.429</td>
<td>8.24E-04</td>
</tr>
<tr>
<td>KEGG</td>
<td>Pathways in cancer</td>
<td>9.131</td>
<td>2.727</td>
<td>0.00109675</td>
</tr>
</tbody>
</table>

**Discussion**

From the regulation network of AS, we could find that many TFs and target genes closely related with AS have been linked by our method. The STAT1 and NFKB1 were hub nodes in our transcriptome network (Figure 1). Some of TFs and target genes have been proved related to AS in previous study, such as STAT1, STAT4, NFKB1, TNFSF10, IL2RA, and IL2RB. Importantly, we also predicted some new TFs and target genes response to AS, such as BXDC5, and EGFRT. Further analysis indicated some significant pathway associated with AS, including antigen processing and presentation, cytokine-cytokine receptor interaction, systemic lupus erythematosus, and apoptosis. We would in detail discuss our findings as following.

NFKB1 is an important transcription regulator that is activated by various intra- and extra-cellular stimuli such as cytokines, ultraviolet irradiation, and viral products. Activated NFKB1 translocates into the nucleus and stimulates the expression of genes involved in a wide variety of biological functions including inflammation and immunity, develop-
mRNA, and serum sTRAIL were significantly elevated in AS patients compared with healthy controls, and there was a close association between TRAIL mRNA and sTRAIL levels. In conclusion, up-regulated expression of TRAIL might be somewhat specific for evaluation of AS\(^\text{27}\). Our results also suggested NF-kappaB might regulate TNFSF10 expression.

STAT1, which is the other hub gene in Figure 1, is a member of the STAT protein family. In response to cytokines and growth factor, such as interferon, STAT family members are phosphorylated by the receptor associated kinases, and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators. STAT1 was up-regulated in ankylosing spondylitis as an IFN-responsive gene\(^\text{13}\). Likewise, STAT4 also plays a central role in IL-12 and IL-23 signaling which associated with AS\(^\text{28}\). IL2RA, IL2RB, and IL2RG constitute the high-affinity IL2 receptor. It has been demonstrated that AS is a chronic inflammatory disease. However, cytokines, such as interleukin-2, are soluble proteins that have specific roles in inflammatory response, arranging the interaction between cells of the immune system both in natural and specific immune reactions. The study results showed serum IL-2R was significantly elevated in patients with AS. And only the sIL-2R level was correlated with Bath AS Metrology Index and Bath AS Functional Index. Therefore, it was suggested that sIL-2R may have a role in the pathogenesis of AS and that their serum levels can be used as disease activity parameters and tools for diagnosis\(^\text{29,30}\).

From the regulation works between TFs and pathways in AS (Figure 2), we could find that many TFs and pathways closely related with AS. The nodes STAT1, STAT4 and NFKB1 are shown as hub nodes in our network. Antigen processing and presentation, cytokine-cytokine receptor interaction pathway, type I diabetes mellitus and T cell receptor signaling pathway, which were significant pathways in AS, were regulated by the STAT1, STAT4 and NFKB1.

Several recent studies have been reported that antigen processing-presentation pathway is associated with AS. HLA-B27 binds unique peptides of microbial or self-origin (antigen) and presents them to CD8+ T cells. In this process, large multifunctional proteases and transporters associated with antigen presentation act as chaperones for peptide transport and have been studied in the context of AS susceptibility. Antigen peptides generated by proteasomal degradation can be fur-

**Figure 1.** Regulation network construction for ankylosing spondylitis. The triangle denotes transcription factor and the circle denotes targeting genes. The red line suggests that the transcription factor could activate their target gene. In contrast, the green line suggests that the transcription factor could inhibit the expression of their target genes.
Analysis of candidate genes

Further subjected to amino peptidase-mediated trimming before reaching the optimal size for HLA Class I binding. Antigen peptides transport into the lumen of the endoplasmic reticulum (regulated by TAP) and are bound to the class I molecule in a process of assisted loading involving multiple proteins collectively known as the peptide-loading complex, including calreticulin, ERP57, and tapasin.31,32

Cytokine-cytokine receptor interaction pathway also has been demonstrated involved in AS, such as TNFRSF10/TRAILR interaction and IL-2/IL-2R interaction. In addition, interleukin-23, as a cytokine, was indicated over-expression in AS. IL-23R, a member of haemopoietin receptor superfamily, is the IL-23R-specific component of the heterodimeric receptor for the cytokine IL-23, which was also identified in AS. This suggests that dysregulation of the IL-23/IL-23R axis (i.e., via IL-23 over-expression and/or altered IL-23R expression or function) might be specifically responsible for the increased occurrence of inflammatory lesions observed in AS.33,34

Although apoptosis pathway was not significant pathway in our analysis results, there was evidence that apoptosis pathway played an important role in AS. Apoptosis pathway, as one of the regulation mechanisms of cell homeostasis, contributed to AS progression. Tumour necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) is a transmembrane (type II) glycoprotein, including membrane-binding TRAIL (mTRAIL) and soluble TRAIL (sTRAIL), which triggers apoptosis through interaction with the death receptors DR4 and DR5. Recently, it has been reported that serum sTRAIL level is up-regulated in AS patients and there are clear associations with CRP, TNF-a, IL-12 and TLR4 mRNA or protein in HLA-B27-positive AS patients, which suggested that sTRAIL involved in apoptosis of synovial fibroblasts.35

In addition, systemic lupus erythematosus (SLE) also has been demonstrated AS complication when treated with some drugs. There was evidence that SLE was drug-induced disease with infliximab in AS patients. A 65 yr-old woman diagnosed with AS according to the modified New York criteria. After high doses of NSAIDs treatment for 3yr failed to control the symptoms, infliximab was advocated. However, forty-eight

![Figure 2. Regulation network of TF-PATHWAY. The triangle denotes transcription factor (TF) and the rectangle denotes the pathway. The red line suggests that the transcription factor could regulate the pathway.](image-url)
hours after the fifth infliximab treatment, the patient was hospitalized with general malaise, fever, myalgias, acute polyarthritis and morning stiffness of more than 1 h. Serological studies showed ANA 1/5120, anti-dsDNA 38.7, negative anti-histone antibodies, CRP 8.8 mg/dl (normal value < 0.8 mg/dl), erythrocyte sedimentation rate 38 mm in the first hour and lymphopenia, that is SLE symptom<sup>14</sup>. Type I diabetes mellitus pathway was suggested as significant pathway in our results, but little was known. We proposed type I diabetes mellitus also may be only some complication in AS patients as SLE.

Conclusions

AS development may attribute to antigen processing and presentation, cytokine-cytokine receptor interaction, and apoptosis pathway. AS for systemic lupus erythematosus, and type I diabetes mellitus, we proposed they all may be only some complication in AS patients when treated with drug.

A deeper understanding of transcription factors, their target genes and pathways remain an area of intense research activity in futures. Our regulation network is useful in investigating the complex interacting mechanisms of transcription factors and their targets in disease. However, further investigations are still needed to confirm the conclusion.

Conflict of Interest

The Authors declare that there are no conflicts of interest.

References

16) WINGENDER E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Briefings in bioinformatics 2008; 9: 326.


