Obtain osteoarthritis related molecular signature genes through regulation network

YAWEI LI, BING WANG, GUOHUA LV, GUANGZHONG XIONG, WEI DONG LIU and LEI LI

Department of Spinal Surgery, The Second Xiangya Hospital of Central South University, Hunan 410011, P.R. China

Received June 14, 2011; Accepted September 7, 2011

DOI: 10.3892/mmr.2011.595

Abstract. Osteoarthritis (OA), also known as degenerative joint disease or osteoarthrosis, is the most common form of arthritis. OA occurs when cartilage in the joints wears down over time. We used the GSE1919 series to identify potential genes that correlated to OA. The aim of our study was to obtain a molecular signature of OA through the regulation network based on differentially expressed genes. From the result of regulation network construction in OA, a number of transcription factors (TFs) and pathways closely related to OA were linked by our method. Peroxisome proliferator-activated receptor $\gamma$ also arises as hub nodes in our transcriptome network and certain TFs containing CEBPD, EGR2 and ETS2 were shown to be related to OA by a previous study.

Introduction

Osteoarthritis (OA) is a type of arthritis caused by the breakdown and eventual loss of the cartilage of one or more joints. OA, also known as degenerative arthritis or degenerative joint disease, is a group of mechanical abnormalities involving degradation of joints, including articular cartilage and subchondral bone (1). OA commonly affects the hands, feet, spine and large weight-bearing joints, such as the hips and knees (2). OA can be defined by symptoms or pathology. The pathology of OA involves the whole joint in a disease process that includes focal and progressive hyaline articular cartilage loss with concomitant changes in the bone below the cartilage, including development of marginal outgrowths, osteophytes and increased thickness of the bony envelope (bony sclerosis). DNA microarray analysis as a global approach is applied to investigate physiological mechanisms in health and disease (3). The development of reliable gene expression profiling technology has an increasing impact on our understanding of OA (4). The differentially expressed genes (DEGs) found through the expression profiles may play significant roles in OA and potentially serve as biomarkers in the diagnosis and prognosis of OA (5).

Such an approach should be established and potential targets for combating the disease should be identified. We suggest that a transcriptome network should be developed in order that a set of transcription factors (TFs), regulating the DEGs induced by OA, can be identified and modulated to this network. Further analysis of the genes and pathways in the network may identify potential mechanisms that respond to OA. The present study does not address this regulation network, but aims to identify the significant pathways associated with OA.

Materials and methods

Data source

Affymetrix microarray data. The transcription profiles of GSE1919 (6) were obtained from a public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/), which is based on the Affymetrix GPL91 platform data. Rheumatoid arthritis expression profiles were collected.

Pathway data. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of online databases dealing with genomes, enzymatic pathways and biological chemicals (7). The PATHWAY database records networks of molecular interactions in the cells and variants thereof that are specific to particular organisms (http://www.genome.jp/kegg/). A total of 130 pathways, involving 2,287 genes, were collected from KEGG.

Regulation data. There are approximately 2,600 proteins in the human genome that contain DNA-binding domains and the majority of these are presumed to function as TFs (8). The combined use of a subset of the approximately 2,000 human TFs easily accounts for the unique regulation of each gene in the human genome during development (9).

These TFs are grouped into five super class families, based on the presence of conserved DNA-binding domains. The TRANSFAC database contains data on TFs, their experimentally proven binding sites and regulated genes (10).

The Transcriptional Regulatory Element Database (TRED) was constructed in response to the increasing requirement for an integrated repository for cis- and trans- regulatory elements in mammals (11). TRED undertook the curation for...
transcriptional regulation information, including TF-binding motifs and experimental evidence. The curation is currently focusing on target genes of 36 cancer-related TF families. A total of 774 pairs of regulatory relationships between 219 TFs and 265 target genes were obtained from TRANSFAC (http://www.gene-regulation.com/pub/databases.html).

A total of 5,772 pairs of regulatory relationships between 102 TFs and 2,920 target genes were obtained from TRED (http://rulai.cshl.edu/TRED).

Combining the two regulation datasets, a total of 6,328 regulatory relationships between 276 TFs and 3,002 target genes were obtained (Table I).

Analysis of differentially expressed genes (DEGs). For the GSE1919 dataset, the limma method (12) was used to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the RMA method, with the default settings implemented in Bioconductor, and then the linear model was constructed. Only DEGs with a fold change value $>2$ and $P<0.05$ were selected.

Co-expression analysis. To demonstrate the potential regulatory relationships, the Pearson’s correlation coefficient (PCC) was calculated for all pair-wise comparisons of the gene-expression values between TFs and the DEGs. The regulatory relationships, whose absolute PCC was $>0.6$, were considered to be significant.

Gene ontology analysis. The BiNGO analysis (13) was used to identify over-represented gene ontology (GO) categories in the biological process.

Regulation network construction. Using the regulation data that had been collected from the TRANSFAC and TRED databases, we matched the relationships between differentially expressed TFs and their differentially expressed target genes.

Based on the above two regulation datasets and the pathway relationships of the target genes, we constructed the regulation networks using Cytoscape (14). Based on the significant relationships (PCC $>0.6$ or PCC $<-0.6$) between TFs and their target genes, 45 putative regulatory relationships were predicted between 20 TFs and 29 target genes.

Significance analysis of pathway. We adopted an impact analysis that included the statistical significance of the set of pathway genes but also considered other crucial factors, such as the magnitude of each gene's expression change, the topology of the signaling pathway and their interactions (15). In this model, the impact factor (IF) of a pathway $\Pi_i$ was calculated as the sum of two terms:

$$\text{IF} (\Pi_i) = \log \left( \frac{1}{p_i} \right) + \sum_{g \in \Pi_i} \Delta E (g)$$

The first term is a probabilistic term that captures the significance of the given pathway $P_i$ from the perspective of the set of genes contained in it. This term is obtained by using the hyper geometric model, in which $P_i$ is the probability of obtaining at least the observed number of DEG, $N_{de}(\Pi_i)$, by chance (16,17).

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 $g$ on the given pathway $P_i$. The PF of a gene $g$ is calculated as:

$$\text{PF} (g) = \Delta E (g) + \sum_{u \in \Pi_i} \beta_{ug} \frac{PF (u)}{\Delta E (g)}$$

In this equation, the first term $\Delta E (g)$ captures the quantitative information measured in the gene expression experiment. The factor $\Delta E (g)$ is the normalized measured expression change of the gene $g$. The first term $\Delta E (g)$ 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 provides this information regarding the type of interaction of two genes in the description of pathway topology). $\Delta E (g)$ is the set of all such genes upstream of $g$. Normalization was achieved with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, $N_{de}(P_i)$. In order to render the IFs as independent from the technology as possible and also comparable between problems, we also divided the second term in equation 1 by the mean absolute fold change $\Delta E$, calculated across all differentially expressed genes. The result of the significance analysis of the pathway is shown in Table III.

Regulation network construction in osteoarthritis. To identify pathway-related DEGs of OA, we obtained publicly available

<table>
<thead>
<tr>
<th>Source</th>
<th>Regulation</th>
<th>TFs</th>
<th>Targets</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRED</td>
<td>5,722</td>
<td>102</td>
<td>2,920</td>
<td><a href="http://rulai.cshl.edu/TRED/">http://rulai.cshl.edu/TRED/</a></td>
</tr>
<tr>
<td>Total</td>
<td>6,328</td>
<td>276</td>
<td>3,002</td>
<td></td>
</tr>
</tbody>
</table>
Following the microarray analysis, the DEGs with a fold change value >2 of GSE1919 and P<0.05 were selected. A total of 888 genes were selected as DEGs from GSE1919. To obtain the regulatory relationships, the co-expressed value (PCC ≥0.6) was selected as the threshold. We then obtained 14 regulatory relationships between 4 differentially expressed TFs and their 12 differentially expressed target genes. By integrating the regulatory relationships above, a regulation network of OA was constructed between TFs and their target genes (Fig. 1).

In this network, CEBPD and peroxisome proliferator-activated receptor γ (PPARG) with higher degrees from a local network indicate that these genes play a significant role in OA. Additionally, TF cascades CEBPD and PPARG, which regulate the IL6 target gene, were observed in this network.

Gene ontology analysis of the regulation network in osteoarthritis. A number of GO categories were enriched among these genes in the regulatory network, including response to endogenous stimulus, response to organic substance, homeostatic process and regulation of biological quality (Table II).

Significant pathway in osteoarthritis. To identify the relevant pathways altered in heart failure, we used a statistical approach on the pathway levels. Significance analysis at the single gene level may suffer from the limited number of samples and experimental noise that can severely limit the power of the selected statistical test. The pathway provides an alternative method with which to relax the significance threshold applied to single genes and may lead to a more accurate biological interpretation. Therefore, we adopted a pathway-based impact analysis method that contained numerous factors, including the statistical significance of the set of DEGs in the pathway, the magnitude of each gene's expression change, the topology of the signaling pathway and their interactions. The impact analysis method yielded a number of significant pathways including the PPAR signaling pathway, antigen processing and presentation, and cytokine-cytokine receptor interaction (Table III).

Discussion

From the result of regulation network construction in OA, we have found that numerous TFs and pathways that are closely correlated to OA were linked by our method. TF PPARG also arose as hub nodes in our transcriptome network, and certain TFs containing CEBPD, early growth response 2 (EGR2) and ETS2 were shown to be associated with OA in a previous study (18).

PPARG encodes a member of the PPAR subfamily of nuclear receptors. PPARs form heterodimers with retinoid X receptors and regulate the transcription of various genes. PPARG is a nuclear receptor involved in the regulation of various cell processes. Previous publications (19) have shown...
Table III. Significant pathway analysis.

<table>
<thead>
<tr>
<th>Database name</th>
<th>Pathway name</th>
<th>Impact factor</th>
<th>Pathway genes in input (%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG</td>
<td>PPAR signaling pathway</td>
<td>16.496</td>
<td>14.286</td>
<td>9.05E-08</td>
</tr>
<tr>
<td>KEGG</td>
<td>Antigen processing and presentation</td>
<td>16.267</td>
<td>4.494</td>
<td>0.042459</td>
</tr>
<tr>
<td>KEGG</td>
<td>Cytokine-cytokine receptor interaction</td>
<td>12.575</td>
<td>4.183</td>
<td>0.002625</td>
</tr>
<tr>
<td>KEGG</td>
<td>Toll-like receptor signaling pathway</td>
<td>11.014</td>
<td>4.902</td>
<td>0.028094</td>
</tr>
<tr>
<td>KEGG</td>
<td>Graft vs. host disease</td>
<td>8.934</td>
<td>9.524</td>
<td>0.002333</td>
</tr>
<tr>
<td>KEGG</td>
<td>Bladder cancer</td>
<td>8.577</td>
<td>11.905</td>
<td>0.001170</td>
</tr>
<tr>
<td>KEGG</td>
<td>Ribosome</td>
<td>8.098</td>
<td>0.990</td>
<td>0.757130</td>
</tr>
<tr>
<td>KEGG</td>
<td>Leukocyte transendothelial migration</td>
<td>7.058</td>
<td>2.521</td>
<td>0.281523</td>
</tr>
<tr>
<td>KEGG</td>
<td>Hematopoietic cell lineage</td>
<td>6.113</td>
<td>5.747</td>
<td>0.021378</td>
</tr>
<tr>
<td>KEGG</td>
<td>Pathways in cancer</td>
<td>5.285</td>
<td>3.030</td>
<td>0.051283</td>
</tr>
</tbody>
</table>

Figure 2. Regulation network of TF PATHWAY. The triangles denote transcription factor (TF) and diamonds denote pathways. The arrowed line indicates that TFs activate the pathways in osteoarthritis (OA). By contrast, the bracketed line indicates that TFs inhibit the expression of the pathways in OA.
that PPARG activators exhibit anti-inflammatory and chondroprotective properties in vitro. These activators also enhance the clinical course and histopathological features in an experimental animal model with OA (20). Pioglitazone, a PPARG agonist, was capable of reducing the severity of experimental OA in pigs. This effect was associated with a reduction in the levels of matrix metalloproteinase (MMP)-13 and interleukin (IL)-1β, which are known to play a crucial role in the pathophysiology of OA lesions (21).

CEBPα encodes a bZIP TF, which binds to certain DNA regulatory regions. It is crucial to the regulation of genes involved in immune and inflammatory responses and may be involved in the regulation of genes associated with the activation and differentiation of macrophages (22). C/EBP TFs are involved in the activation of OA cells, and endoplasmic reticulum stress causes cell death in OA (23).

EGR2 encodes a TF with three tandem C2H2-type zinc fingers. Egr-2 is an intrinsic regulator of effector T cells and controls the expansion of self-reactive T cells and the development of autoimmune disease (24). However, to the best of our knowledge, no publication has mentioned that EGR2 is involved in atherosclerosis.

ETS2 regulates numerous genes and is involved in stem-cell development, cell senescence and death, and tumorigenesis. The genetic programs regulated by Ets2 significantly affect the development and differentiation of osteoblasts. Additionally, Ets2 overexpression in transgenic is responsible for the genesis of the same type of skeletal abnormalities that are found in Down's syndrome (25,26).

TFs bind to specific genes to control the flow of genetic information from DNA to mRNA. From the result of the regulation network between TFs and their target genes in OA we have found that there are a number of target genes that may be associated with OA.

IL-6 encodes a cytokine that functions in inflammation and the maturation of B cells. The functioning of this gene is involved in a wide variety of inflammation-associated diseases, including susceptibility to diabetes mellitus and systemic juvenile rheumatoid arthritis (27,28). A common feature of OA is the alteration of type II collagen, which is composed of α1(II) chains. IL-6 inhibits COL2A1 transcription encoded by the COL2A1 gene by decreasing the Sp1 Sp3 ratio and DNA-binding activities (29).

CDKN1A (p21) is a cyclin-dependent kinase inhibitor. p21 binds to and inhibits the activity of cyclin-CDK2 or -CDK4 complexes, and thus acts as a regulator of cell-cycle progression at G1. Cell cycle progression inhibitor p21 is expressed in normal and is significantly down-regulated in OA articular chondrocytes, which may mediate the re-initiation of cell proliferation in OA cartilage (30).

MMP-9 belongs to the MMP family, which is involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction and tissue remodeling, as well as in disease processes, such as arthritis and metastasis (31). MMP-9 expression is increased in OA cartilage relative to normal cartilage. In addition, it is also increased in OA bone relative to normal bone according to an investigation of animals with OA (32). Studies have shown that HIF-2α directly induces the expression of MMP-9 to affect the OA process by causing cartilage destruction (33).

CD4 gene encodes a membrane glycoprotein of T lymphocytes that interacts with major histocompatibility complex class II antigens and is also a receptor for the human immunodeficiency virus. CD4 protein functions to initiate or augment the early phase of T-cell activation. CD4(+) T cells promote disease progression of OA by inducing macrophage inflammatory protein-1γ. CD4(+) T cells (express the CD4 protein on their surface) are activated during the onset of OA and cause marked damage to cartilage at a later stage (34).

From the result of regulation network between TFs and pathways in OA, we found that there are numerous pathways, such as the PPAR signaling pathway, antigen processing and presentation and cytokine-cytokine receptor interaction, which are closely correlated to OA and have been linked by our method.

The PPAR signaling pathway has been proven to be associated with OA. Microsomal prostaglandin E synthase-1 (mPGES-1) and its functional coupled protein cyclooxygenase-2 (COX-2) are inflammatory genes, both of which are overexpressed in OA patients and regulated by PPRG (35,36). The MMP family plays a significant role in the process of OA, denoting cartilage destruction. A number of subtypes of this family are increased in OA, while PPRG down-regulates their expression (37). Those investigations indicated that PPRG may be a negative modulator of OA.

Evidence has indicated that antigen processing and presentation pathway may be associated with OA. Cathepsins play a vital role in antigen processing and presentation; cathepsin S and K expression levels increase following stimulation of RA- and OA-derived synovial fibroblasts with IL-1β and tumor necrosis factor-α (38). Interferon-γ (IFN-γ) is a cytokine that is critical for innate and adaptive immunity against viral and intracellular bacterial infections. T cells from OA synovial fluids express an activated phenotype and synthesized IFN-γ, supporting the hypothesis that inflammatory cells play a significant role in OA disease progression and chronicity (39).

The Cytokine-cytokine receptor interaction pathway was also demonstrated to be associated with OA. Transforming growth factor-β (TGF-β) is a protein that controls proliferation, cellular differentiation and other functions in the majority of cells. A recent publication revealed that TGF-β2 and TGF-β3 protein levels are significantly higher and are positively correlated with pain in OA patients (38). Cytokine of IL-1β belongs to IL-1 family of ligands and receptors, which is associated with acute and chronic inflammation. IL-1β has emerged as a therapeutic target for an expanding number of systemic and local inflammatory conditions termed ‘auto-inflammatory’ diseases. In these diseases, including OA, neutralization of IL-1β results in a rapid and sustained reduction in disease severity (41). Suppressor of cytokine signaling (SOCS) proteins are inhibitors of cytokine signaling that function via the JAK/STAT pathway. The expression of SOCS2 and CIS-1 (two members of SOCS) in OA proteins is attenuated, indicating a role for SOCS in the pathology of OA (42).

The toll-like receptor (TLR) signaling pathway may play a significant role in OA. TLRs are a class of proteins that play a key role in the innate immune system. They are single, membrane-spanning, non-catalytic receptors that recognize
structurally conserved molecules derived from microbes (43). TLR-4 and TLR-3 are highly upregulated in OA patients (44). TLR ligands mediate the catabolic pathways leading to a significant increase in the release of proteoglycan and type II collagen degradation product in human OA chondrocytes. Activation of chondrocyte TLRs leads to differential collagen gene activation. According to the analysis results of the femoral heads, treatment of chondrocytes with TLR1/2 or TLR6/2 ligands resulted in collagen resorption (45). TLR signal pathway is a possible mechanism in warming-needle therapy, an effective therapy for knee OA with deficiency-cold syndrome (46).

In conclusion, sets of key TFs and their target genes and pathways potentially associated with OA were identified in the present study. Although previous employment of these methods has been somewhat effective, further investigations are required to confirm our findings.

References