Joint Bone Spine 83 (2016) 545–551

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

Original article

Identification of upstream regulators for synovial expression signature genes in osteoarthritis

Qi Fei , JiSheng Lin , Hai Meng , BingQiang Wang , Yong Yang , Qi Wang , Nan Su , Jinjun Li , Dong Li *

Department of Orthopaedics, Beijing Friendship Hospital, Capital Medical University, 95, Yong'an Road, Beijing 100050, China

ARTICLE INFO

Article history:
Accepted 14 September 2015
Available online 29 January 2016

Keywords:
Integrated analysis
Gene expression data
Osteoarthritis
Microarray
Differentially expressed genes

ABSTRACT

Objectives: The detection of transcription factors (TFs) for OA signature genes provides better clues to the underlying regulatory mechanisms and therapeutic applications.
Methods: We searched GEO database for synovial expression profiling from different OA microarray studies to perform a systematic analysis. Functional annotation of DEGs was conducted, including gene ontology (GO) enrichment analysis and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis. Based on motif databases and the results from integrated analysis of current gene expression data, a global transcriptional regulatory network was constructed, and the upstream TFs were identified for OA signature genes.
Results: Six GEO datasets were obtained. Totally, 805 genes across the studies were consistently differentially expressed in OA (469 up-regulated and 336 down-regulated genes) with FDR $\leq$ 0.01. Supporting an involvement of ECM in the development of OA, we showed that ECM-receptor interaction was the most significant pathway in our KEGG analysis (P = 5.92E-12). Sixty-one differentially expressed TFs were identified with FDR $\leq$ 0.05. The constructed OA-specific regulatory networks consisted of 648 TF-target interactions between 51 TFs and 429 DEGs in the context of OA. The top 10 TFs covering the most downstream DEGs were identified as crucial TFs involved in the development of OA, including ARID3A, NFIC, ZNF354C, NR4A2, BRCA1, EHF, FOXL1, FOXC1, EGR1, and HOXA5.
Conclusion: This integrated analysis has identified the OA signature, providing clues to pathogenesis of OA at the molecular level, which may be also used as diagnostic markers for OA. Some crucial upstream regulators, such as NR4A2, EHF, and EGR1 may be considered as potential new therapeutic targets for OA.

1. Introduction

Osteoarthritis (OA) is the most prevalent joint disease characterized by progressive degradation of the articular cartilage. Regarded as a disease of age-related wear and tear on joints of the hand, knee or hip, OA is very common in older adults but relatively rare in younger adults. In addition to cartilage loss, OA also involves pathological changes in the structure and function of ligaments, capsular tissue, periarticular muscle, nerve, subchondral bone, meniscus and synovium [1,2].

Despite considerable research efforts in the past few decades, the pathogenesis of OA remains far from being fully understood. Nevertheless, it is clear that there is a strong genetic component in the etiology of OA [3]. In the chondrocytes of OA patients, the expression profile of genes encoding various types of protease [4,5], proinflammatory cytokines, cell receptors, matrix proteins (types II and X collagens and aggrecan) and several transcription factors [6] is altered. Such alteration leads to destroy of the balance between anabolism and catabolism in favor of catabolism and thus further aggravates the degradative phenotype.

Recently, various high-throughput technologies, making it possible to screen the whole genome, have revolutionized the research on unusual genomic alterations in human diseases. Particularly, the use of microarrays has significantly advanced the knowledge about the cellular and molecular events occurring in OA [7–13]. Although the newly emerged technology of RNA sequencing (RNA seq) has advantages over microarrays, it has not yet been used in the OA field. To date, the information on differentially expressed genes (DEGs) in the pathogenesis of OA is obtained primarily from studies employing microarrays [14]. Owing to considerable differences in

* Corresponding author.
E-mail address: spinefei@163.com (D. Li).

http://dx.doi.org/10.1016/j.jbspin.2015.09.001
1297-319X/

sample size, laboratory protocols, microarray platforms and analysis techniques among different microarray studies, inconsistent results have been obtained [15]. Capable of increasing the statistical power, integrated analysis of gene expression data from multiple sources has emerged as a novel approach for detecting DEGs and assessing the heterogeneity of individual microarray data in a more accurate manner than individual microarray study [16,17].

Further, this strategy provides a chance to develop biomarkers for diagnosis, which is highly desirable for practical usage in terms of prediction of prognosis and drug response. Sweeney et al. performed a time-course–based integrated analysis of gene expression data in sepsis, and 11 DEGs was identified to distinguish sterile inflammation from infectious inflammation with excellent diagnostic power [18]. Oh et al. applied a genome-wide gene expression of 177 colorectal patients in the GEO database from different microarray studies, and obtained 85 prognostic gene expression signature genes, which could accurately discriminate colorectal cancer patients with good prognosis from those with prognosis [19].

In this study, we also performed a systematic analysis of current OA gene expression data to identify DEGs in synovial tissues between OA and normal control (NC), which may be considered their utility as diagnostic markers. Transcription factors (TFs) could enhance or inhibit gene transcription via binding to specific DNA sequences generally located in the promoter region of genes. Taking advantage of the resource of motif databases such as TRANSFAC and the results from integrated analysis of current gene expression data, we identified a set of TFs mediating gene expression in the process of OA pathogenesis, and constructed OA-specific transcriptional regulatory networks for a systematic understanding of disease progression at the molecular level. Hopefully, identification of crucial upstream regulators would provide clues to potential new therapeutic targets for the disease.

2. Methods

2.1. Identification of eligible OA gene expression datasets

Considering that OA was characterized by the breakdown of articular cartilage in synovial joints, we selected synovial gene expression profiling studies of OA on the Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo) [20]. Those datasets that were obtained from microarray experiments on the gene expression of synovial tissues in OA and NC, were downloaded.

2.2. Statistical analysis

The raw microarray datasets were downloaded, and preprocessed with log2 transformation and Z-score normalization for each study. The Linear Models for Microarray Data (Limma) package in R was used to identify the differently expressed genes between OA and controls by two-tailed Student's *t*-test, and *P*-value was obtained. Further false discovery rate (FDR) was calculated for multiple comparisons using the Benjamini & Hochberg method. Genes with FDR < 0.01 were considered as differently expressed genes (DEGs). Hierarchical clustering analysis was performed using the "heatmap.2" function of the R/Bioconductor package "gplots" [21].

2.3. Functional annotation of DEGs

The biological functions of the DEGs in the pathogenesis of OA were interpreted by gene ontology (GO) enrichment analysis by GO-rilla, which is a web-accessible program for GO enrichment analysis [22]. GO provides functional annotation and classification for analyzing the gene sets data (i.e., biological process, cellular component, and molecular function). In order to understand the biological pathways that the DEGs were involved in, the Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis was also used. The web-based software GENECODIS [23] was used to perform pathway enrichment analysis, and it was considered to be statistically significant when a threshold *P*-value < 0.05 was used in the hypergeometric test.

2.4. Construction of OA-specific transcriptional regulatory networks

TFs act as drivers or master regulators of gene expression to provide better clues to the underlying regulatory mechanisms. To provide a comprehensive and deeper knowledge about gene regulation underling OA, we extract information about TFs likely involved in regulating these DEGs. Based on OA signatures generated from integrated analysis, we searched TRANSFAC which is a database of TFs, their genomic binding sites, and DNA binding site sequence profiles for DEG coded TFs and their targeted genes, and used TRANSFAC position weight matrix (PWM) for gene promoter scanning [24] to identify DEGs which has the binding site of the TF in the promoter region. The transcriptional regulatory networks were visualized by Cytoscape [25].

3. Results

General information on the included studies: a total of 6 expression-profiling studies met the inclusion criteria and were included. The general information of these studies is detailed in Table 1. Of these studies, 76 cases of OA and 50 cases of controls were contained.

DEGs in OA: a total of 18,625 genes were assessed across 6 expression profiling studies. Eight hundred and five were differentially expressed in OA with FDR ≤ 0.01 as compared with the normal control. Among these DEGs, 469 were up-regulated and 336 were down-regulated. Listed in Table 2 is a list of the top 10 most significantly up- or down-regulated genes. The full list of these genes is displayed in the Supplementary data, Table S1 [see the supplementary material associated with this article online]. We performed a hierarchical clustering analysis on the OA cohort. As expected, all samples were classified into normal and OA group based on this signature (Fig. 1).

Annotated functions of DEGs: P-values for the three GO categories (biological process, molecular function and cellular component) were calculated and analyzed by GO-rilla. Genes with a nominal significance level of $P < 0.001$ were selected and tested against the background set of all genes with GO annotations. The significantly enriched GO terms for biological process were small molecule metabolic process (GO: 0044281, $P = 2.86E\text{-}15$) and lipid metabolic process (GO: 0006629, $P = 2.23E\text{-}12$), while those for molecular function were cofactor binding (GO: 0048037, $P = 8.65E\text{-}10$) and coenzyme binding (GO: 0050662, $P = 9.60E\text{-}10$). And those for cellular component were extracellular organelle (GO: 0043230, $P = 7.70E\text{-}13$) and extracellular membrane-bounded organelle (GO: 0065010, $P = 7.70E\text{-}13$) (Supplementary data, Figs. S1–S3).

Summarized in Table 3 are top 15 the most significantly enriched signaling pathways of the identified DEGs. Supporting an involvement of ECM in the development of OA, we showed that ECM-receptor interaction was the most significant pathway in our KEGG analysis ($P = 5.92E\text{-}12$). And cell cycle ($P = 7.43E\text{-}11$) and fatty acid metabolism ($P = 1.09E\text{-}10$) were also highly enriched.

Transcriptional regulatory networks: undirected and directed regulatory networks with DEG – coded TFs and regulated DEGs were created. Based on TRANSFAC, 61 differentially expressed TFs were identified with FDR ≤ 0.05 (Supplementary data, Table S2).