Identification and function analysis of contrary genes in Dupuytren's contracture

XIANGLU JI, FENG TIAN and LIJIE TIAN

Department of Hand and Foot Surgery, Shengjing Hospital, China Medical University, Shenyang, Liaoning 110003, P.R. China

Received May 16, 2014; Accepted January 29, 2015

DOI: 10.3892/mmr.2015.3458

Abstract. The present study aimed to analyze the expression of genes involved in Dupuytren's contracture (DC), using bioinformatic methods. The profile of GSE21221 was downloaded from the gene expression omnibus, which included six samples, derived from fibroblasts and six healthy control samples, derived from carpal-tunnel fibroblasts. A Distributed Intrusion Detection System was used in order to identify differentially expressed genes. The term contrary genes is proposed. Contrary genes were the genes that exhibited opposite expression patterns in the positive and negative groups, and likely exhibited opposite functions. These were identified using Coexpress software. Gene ontology (GO) function analysis was conducted for the contrary genes. A network of GO terms was constructed using the reduce and visualize gene ontology database. Significantly expressed genes (801) and contrary genes (98) were screened. A significant association was observed between Chitinase-3-like protein 1 and ten genes in the positive gene set. Positive regulation of transcription and the activation of nuclear factor-κB (NF-κB)-inducing kinase activity exhibited the highest degree values in the network of GO terms. In the present study, the expression of genes involved in the development of DC was analyzed, and the concept of contrary genes proposed. The genes identified in the present study are involved in the positive regulation of transcription and activation of NF-κB-inducing kinase activity. The contrary genes and GO terms identified in the present study may potentially be used for DC diagnosis and treatment.

Introduction

Dupuytren's contracture (DC) is a progressive tissue disorder that affects the palmar fascia, and results in digital flexion contracture (1-3). DC incidence ranges from 3-40% worldwide and primarily affects patients >50 years (4). Therefore, it is necessary to investigate novel diagnostic and therapeutic approaches for patients with DC.

A number of treatments for DC have been evaluated, such as radiotherapy, topical vitamin A and E application, and dimethyl sulfoxide injection. However, these approaches have proven unsuitable or ineffective for the clinical treatment of DC (5). More recently, research has focused on potential treatments targeting the molecular processes underlying DC (6,7). Fibrogenic cytokines, which may be capable of inducing the growth of fibroblasts, are involved in the molecular mechanisms underlying DC (8). Satish et al (9) demonstrated that the expression of type XV collagen α 1 chain, proteoglycan 4 and fibrin-1, which are components of the extracellular matrix, is downregulated in DC fibroblasts compared with controls. The following miRNAs were found to be associated with the β-catenin pathway in DC tissue samples: Zinc finger of the cerebellum 1, wingless-type mouse mammary tumor virus integration site 5A and transforming growth factor β 1, (10). Musculoaponeurotic fibrosarcoma oncogene homolog B has been reported to be expressed in DC samples and not in healthy samples (11). Furthermore, a number of matrix metalloproteinases (MMPs), including MMP 13, MMP 16 and MMP 19 are expressed in DC samples but not in healthy samples (12). However, the pathogenesis of DC remains poorly understood. It is therefore necessary to investigate the pathology of DC in order to develop diagnostic and therapeutic approaches to DC.

In the present study, statistical analysis was performed, in order to develop gene expression profiles for samples from patients with DC and from healthy controls. The significantly differentially expressed genes (DEGs) between DC patients and healthy controls were measured. Hierarchical clustering analysis was performed for the DEGs. Antagonistic contrary genes, which exhibited opposite expression patterns in the positive and negative groups, and likely exhibited opposite functions, were then identified. A gene ontology (GO) function analysis was conducted and a network of GO terms was constructed.

Materials and methods

DEG identification. The GSE21221 expression profile was downloaded from the gene expression omnibus (GEO) database.
This included six samples from DC-derived fibroblasts and six control samples from carpal-tunnel-derived fibroblasts, and was conducted using GPL2507 Sentrix® Human-6 Expression BeadChip and GPL10301 GE Healthcare CodeLink Human Whole Genome Bioarray (9). DEGs (P<0.05) were identified using the Distributed Intrusion Detection System (DIDS) (13) with R package version 1.1.

Hierarchical cluster analysis. Hierarchical cluster analysis for DEGs was performed in order to investigate the patterns of DEG classification and co-expression using TreeView (14). Genes whose expression values were present in >80% of samples had a standard deviation (SD) >2 were selected. Expression values were logarithmically standardized using TreeView. The correlations in the similarity matrix were subsequently calculated and a hierarchical cluster analysis was performed.

Identification of contrary genes. A Pearson correlation coefficient was calculated using the co-express software (www.bioinformatics.lu/CoExpress), which analyzes the association between mRNA and miRNA expression. miRNA targets were predicted using TargetScan (http://www.targetscan.org/) and starBase (15). Genes with high expression values <520 and SD value <560 were removed. Between-experiment-normalization was then performed. Genes with a Pearson correlation coefficient >0.9 were defined as positive. Genes that were negatively associated with at least one gene in the positive gene group (Pearson correlation coefficient, <0.9 and correlation coefficient, <0.5 with all the genes in the positive gene group) were defined as negative. Genes in the positive and negative groups exhibited opposite expression patterns. Therefore, in the present study these genes are termed contrary genes, which exhibited opposite expression patterns in the positive and negative groups, and likely exhibited opposite functions.

Function annotation analysis of DEGs. A genecard database (16) was used in order to conduct function analysis of DEGs annotated with GO terms. Cluster analysis was conducted using reduce and visualize gene ontology (REViGO) (17), in order to identify the GO terms (similarity threshold: 0.7). Semantic similarity was measured using the SimRel algorithm (18).

System network analysis. A function network of GO terms was produced using the REViGO database (17), and network visualization was performed using cytoscape version 2.8 (19).
A network analysis plugin (20) was then used in order to analyze the network topology and to identify the significant GO terms in the network.

**Results**

**DEG identification.** The distribution of SD values was used in order to indicate the level of homogeneity in the data. The results suggested that the data exhibited low levels of homogeneity (Fig. 1). A DIDS analysis was performed in order to identify DEGs in the DC samples that had significantly higher expression than normal range, this resulted in the identification of 801 DEGs (P<0.05).

**Hierarchical cluster analysis.** Using the hierarchical clustering method, significant DEGs (775) were identified among the 801 DEGs. Fig. 2 represents a heatmap produced using the cluster software, TreeView. The results suggested that the control and DC samples were clustered in three groups: Clusters one, two and three. In cluster one, genes in the control

### Table I. Significantly correlated genes and correlation coefficients.

<table>
<thead>
<tr>
<th>Node ID1</th>
<th>Node ID2</th>
<th>Correlation coefficient</th>
<th>Node ID1</th>
<th>Node ID2</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHI3L1</td>
<td>EIF2S3</td>
<td>-0.9273</td>
<td>EPHX1</td>
<td>ATF4</td>
<td>-0.90163</td>
</tr>
<tr>
<td>SFRP1</td>
<td>MASP1</td>
<td>0.91333</td>
<td>FAM12A</td>
<td>AKR1C2</td>
<td>0.9186</td>
</tr>
<tr>
<td>SFRP1</td>
<td>FBLN1</td>
<td>0.9048</td>
<td>SAA1</td>
<td>ATF4</td>
<td>-0.9023</td>
</tr>
<tr>
<td>FBLN1</td>
<td>MASP1</td>
<td>0.93677</td>
<td>NFE2L2</td>
<td>AKR1C2</td>
<td>0.94147</td>
</tr>
<tr>
<td>COL15A1</td>
<td>RAB9A</td>
<td>-0.90623</td>
<td>ITM2B</td>
<td>AKR1C2</td>
<td>0.92597</td>
</tr>
<tr>
<td>RND3</td>
<td>AKR1C2</td>
<td>0.92937</td>
<td>RNF19A</td>
<td>ATF4</td>
<td>-0.904</td>
</tr>
<tr>
<td>EIF2S3</td>
<td>AKR1C2</td>
<td>0.92647</td>
<td>IFI16</td>
<td>AKR1C2</td>
<td>0.90557</td>
</tr>
<tr>
<td>DDX3X</td>
<td>AKR1C2</td>
<td>0.9201</td>
<td>NPC1</td>
<td>AKR1C2</td>
<td>0.91037</td>
</tr>
</tbody>
</table>

CHI3L1, chitinase-3-like protein 1; SFRP, secreted frizzled-related protein 1; FBLN, fibulin; COL15A1, collagen type XV α 1; RND3, rho family GTPase 3; EIF2S3, eukaryotic translation initiation factor 2 subunit 3; DDX3X, DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked; MASP1, mannan-binding lectin serine protease 1; CLU, clusterin; RAB, ras-related protein; AKR1C2, aldo-keto reductase family 1, member C2; EPHX1, epoxide hydrolase 1; FAM12A, family with sequence similarity 12, member A; SAA1, serum amyloid A1; NFE2L2, nuclear factor (erythroid-derived 2)-like 2; LAMA, laminin; ITM2B, integral membrane protein 2B; RNF19A, ring finger protein 19A; IFI16, γ-interferon-inducible protein 16; NPC1, Niemann-Pick disease, type C1; ATF4, activating transcription factor 4.

### Table II. Cluster analysis of GO functions.

<table>
<thead>
<tr>
<th>GO ID</th>
<th>GO term</th>
<th>Frequency (%)</th>
<th>Plot size</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster one</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO: 0001568</td>
<td>Blood vessel development</td>
<td>1.36</td>
<td>3.025</td>
<td>0.814</td>
</tr>
<tr>
<td>GO: 0030097</td>
<td>Hemopoiesis</td>
<td>1.46</td>
<td>3.053</td>
<td>0.796</td>
</tr>
<tr>
<td>GO: 0030099</td>
<td>Myeloid cell differentiation</td>
<td>0.64</td>
<td>2.694</td>
<td>0.781</td>
</tr>
<tr>
<td>Cluster two</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO: 0008283</td>
<td>Cell proliferation</td>
<td>4.08</td>
<td>3.501</td>
<td>0.874</td>
</tr>
<tr>
<td>GO: 0060548</td>
<td>Negative regulation of cell death</td>
<td>1.81</td>
<td>3.148</td>
<td>0.674</td>
</tr>
<tr>
<td>GO: 0030307</td>
<td>Positive regulation of cell growth</td>
<td>0.21</td>
<td>2.21</td>
<td>0.722</td>
</tr>
<tr>
<td>GO: 0008284</td>
<td>Positive regulation of cell proliferation</td>
<td>1.65</td>
<td>3.107</td>
<td>0.712</td>
</tr>
<tr>
<td>Cluster three</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO: 0001933</td>
<td>Negative regulation of protein phosphorylation</td>
<td>0.49</td>
<td>2.584</td>
<td>0.666</td>
</tr>
<tr>
<td>GO: 0006486</td>
<td>Protein glycosylation</td>
<td>0.73</td>
<td>2.755</td>
<td>0.787</td>
</tr>
<tr>
<td>GO: 0016567</td>
<td>Protein ubiquitination</td>
<td>1.31</td>
<td>3.007</td>
<td>0.819</td>
</tr>
<tr>
<td>GO: 0006468</td>
<td>Protein phosphorylation</td>
<td>5.22</td>
<td>3.608</td>
<td>0.786</td>
</tr>
</tbody>
</table>

Frequency represents the frequency of the GO terms in the Genecards database and plot size represents the size of GO terms in the integral distribution after correction. Uniqueness represents the average similarity of the GO terms. GO, gene ontology.
samples were upregulated, whereas those in the DC samples were significantly downregulated. In cluster two, genes in the control samples were upregulated and in the DC samples, upregulated and downregulated genes were observed. In cluster three, genes in the control samples were downregulated, whereas those in the DC samples were upregulated. In cluster three, DC and healthy samples were distinct from one another, whereas in clusters one and two, DC and healthy samples were not distinct. Therefore, genes associated with cluster three, may be useful for the development of diagnostic markers associated with DC.

**Identification of contrary genes.** Coexpress software was used in order to conduct a correlation analysis. The average expression value of the 801 DEGs was 523 and the average SD value was 2.24x10^2. Following filtering and standardization of the data, 98 contrary genes were identified. The average expression value of the 98 contrary genes was 3.45x10^3 and
the SD value was 5.56x10^2. Subsequently, 18 significantly correlated contrary gene pairs were identified (Table I). Chitinase-3-like protein 1 (CHI3L1) was negatively correlated with the following ten DEGs in the positive group: DEAD (Asp-Glu-Ala-Asp) box helicase 3; X-Linked (DDX3X); nuclear factor erythroid 2-related factor 2 (NFE2L2); γ-interferon-inducible protein 16 (IFI16); integral membrane protein 2B (ITM2B); Niemann-Pick disease, type C1 (NPC1); rho family GTPase 3 (RND3), family with sequence similarity 129 A (FAM129A); eukaryotic translation initiation factor 2B (ITM2B); laminin, α4 (LAMA4); and aldo-keto reductase family 1, member C2 (AKR1C2; Fig. 3).

Function annotation analysis. A GO function annotation analysis was performed for the CHI3L1 gene and the ten genes with which it was negatively correlated, which resulted in the identification of 132 GO terms. The 132 GO terms were subsequently grouped into three clusters corresponding to the clusters in Fig 2 (Table II). The results suggested that the genes in cluster one are involved in blood vessel development,
hemopoiesis and myeloid cell differentiation. Genes in cluster two are involved in cell proliferation, and the negative and positive regulation of cell death. Genes in cluster three are involved in the negative regulation in protein phosphorylation, glycosylation, ubiquitination and phosphorylation.

**System network analysis.** A system network of the 132 GO terms was constructed (Fig. 4). GO terms exhibiting semantic similarities were removed. This resulted in 86 GO terms and 257 edges, which represent the associations between GO terms. The GO terms, level of positive regulation of transcription and activation of nuclear factor-xB (NF-xB)-inducing kinase activity, exhibited degree levels of 19 and 16, respectively.

**Discussion**

In the present study, a DIDS analysis was conducted, which resulted in the identification of 801 DEGs. A positive and a negative group of DEGs were obtained. Opposing DEGs in the two groups are termed contrary genes in the present study. A network of GO terms was constructed, following a functional analysis, for the CHI3L1 gene and the ten genes with which it was negatively correlated (Table III).

Hierarchical clustering analysis was performed for 775 DEGs, which resulted in three clusters. Differences were observed between the expression patterns of genes in these three clusters. Contrary genes were identified based on the correlation of gene expression level between the DEGs. CHI3L1 expression was found to be negatively correlated with the expression of ten DEGs in the positive group. CHI3L1 is a secreted glycoprotein, which may be termed YKL-40. Forrester et al. (21) performed a genome-wide analysis of exon arrays, which demonstrated that CHI3L1 expression was higher in primary fibroblasts obtained from patients with Dupuytren's disease, compared with those obtained from healthy patients. In the present study, CHI3L1 expression was higher in the DC group than in the control group. Activation of NF-xB-inducing kinase activity exhibited a high value in the GO term network. CHI3L1 has been shown to be induced by inflammatory cytokines, tumor necrosis factor-alpha (TNF-α) and interleukin-1, in articular chondrocytes, and this induction requires sustained activation of NF-xB (22). Furthermore, it has been reported that the expression of TNF-α and interleukin-1 was significantly higher in patients with DC compared with that in controls (23). Therefore, CHI3L1 may be involved in the pathology of DC, via the induction of NF-xB kinase activity. However, the involvement of CHI3L1 in the development of DC requires further analysis. In the present study, the following ten genes were negatively correlated with CHI3L1: DD3X3, NFE2L2, IFI16, ITM2B, NPC1, RND3, FAM129A, EIF2S3, LAMA4 and AKR1C2. To the best of our knowledge, the present study is the first to demonstrate a potential association between these ten genes and DC development. DD3X (24), NFE2L2 (25), IFI16 (26), NPC1 (27), RND3 (28) and AKR1C2 (29) have been found to be positively correlated with inflammation. Therefore, these genes may be associated with DC development. However, the association between these genes and CHI3L1, and their involvement in DC, require further investigation, in order to understand the molecular mechanisms underlying DC development.

Positive regulation of transcription demonstrated the highest degree in the GO term network. The transcription factor gene, ZF9 is associated with DC pathogenesis (30). The expression of the contractile phenotype in Dupuytren's nodular cells is determined by the post-transcriptional regulation of α-smooth muscle actin (31). Therefore, the regulation of transcription may be involved in the pathogenesis of DC.

In the present study, the expression of genes involved in DC was analyzed, and the term, contrary genes, was proposed. The expression of CHI3L1 was shown to be negatively correlated with that of ten other genes. These genes are involved in the positive regulation of transcription and in the activation of NF-xB-inducing kinase activity. Therefore, the results of the present study may be of use in the development of diagnostic markers and treatment for DC.

**References**


