Analysis of crucial molecules involved in herniated discs and degenerative disc disease

Zhigang Qu, Weiwei Miao, Qi Zhang, Zhenyu Wang, Changfeng Fu, Jinhua Han, Yi Liu

The Second Hospital of Jilin University, Department of Orthopedics, Changchun, Jilin Province, China.

OBJECTIVES: Herniated discs and degenerative disc disease are major health problems worldwide. However, their pathogenesis remains obscure. This study aimed to explore the molecular mechanisms of these ailments and to identify underlying therapeutic targets.

MATERIAL AND METHODS: Using the GSE23130 microarray datasets downloaded from the Gene Expression Omnibus database, differentially co-expressed genes and links were identified using the differentially co-expressed gene and link method with a false discovery rate <0.25 as a significant threshold. Subsequently, the underlying molecular mechanisms of the differential co-expression of these genes were investigated using Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis. In addition, the transcriptional regulatory relationship was also investigated.

RESULTS: Through the analysis of the gene expression profiles of different specimens from patients with these diseases, 539 differentially co-expressed genes were identified for these ailments. The ten most significant signaling pathways involving the differentially co-expressed genes were identified by enrichment analysis. Among these pathways, apoptosis and extracellular matrix-receptor interaction pathways have been reported to be related to these diseases. A total of 62 pairs of regulatory relationships between transcription factors and their target genes were identified as critical for the pathogenesis of these diseases.

CONCLUSION: The results of our study will help to identify the mechanisms responsible for herniated discs and degenerative disc disease and provides a theoretical basis for further therapeutic study.

KEYWORDS: Herniated Discs; Degenerative Disc Disease; Transcription Factor; Target Genes; Pathway.

■ INTRODUCTION

The high prevalence of low back pain (LBP) makes disc degeneration a leading health problem. Currently, 70-80% of the adult population reports at least one episode of LBP during their life (1). In the American population, the incidence of symptomatic lumbar disc herniation is 1-2%, and approximately 200,000 lumbar discectomies are performed annually (2). Although there are several causes of spinal pain, intervertebral discs (IVDs) appear to be the most common source of chronic or recurring axial LBP (3). IVD degeneration is defined as an aberrant cell-mediated response to progressive structural failure and is a multi-factorial process influenced by genetics, lifestyle, and biomechanics (4).

Normal adult IVDs are made up of avascular tissue composed of a large amount of extracellular matrix (ECM) with a low cell density (approximately 1% of the total volume). Small solutes, such as glucose, lactic acid, and oxygen, in the inner parts of both the annulus fibrosus and nucleus pulposus are mainly transported by diffusion. The disc cells are responsible for maintaining the integrity of the ECM (5). Changes in the metabolic balance of IVD cells affect the quality and quantity of the ECM and its functional properties, and these changes can therefore be related to disc degeneration (6). Microscopically, disc degeneration is characterized by cell proliferation, cell necrosis or apoptosis, mucous degeneration, granular changes, and ingrowths of nerves or blood vessels (4,6). Cell proliferation leads to cluster formation, particularly in the nucleus pulposus (7). A recent study indicated that genetics plays a crucial role in disc degeneration, explaining 29-54% of the variance in adult populations (8). Environmental factors, such as mechanical loading and smoking, account for the remaining variance, although not all factors have been identified.

In this study, the gene expression profiles of cells from herniated discs and discs with signs of degenerative disc disease, as well as data on transcription factors (TFs) and

target genes, were analyzed to identify underlying signaling pathways and critical molecules that regulate these pathways at the transcriptional level. Recently, a number of methods, such as comparison and cluster analysis, have been adopted to analyze gene expression. However, many studies have reported that the identification of differentially expressed genes using these analyses cannot reveal the connections among various genes, and therefore, it is important to obtain clusters of genes that are differentially co-expressed, share common attributes and function collaboratively (9,10). Using a variety of analytical methods, differentially co-expressed genes (DCGs) in cells from herniated discs and from discs exhibiting signs of degenerative disc disease were analyzed using the DCGL package (11,12) in R.

■ MATERIALS AND METHODS

Microarray analysis

The transcription profile of GSE23130 was obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). Based on the GPL1352 [U133_X3P] Affymetrix Human X3P Array, the RNA expression profiles of IVD cells from 23 patients with herniated discs or degenerative disc disease were analyzed. The 23 tissue samples for the microarray study consisted of one grade I, five grade II, nine grade III, five grade IV, and three grade V herniated discs or degenerative disc disease. All tissue samples were obtained from the Cancer Institute Cooperative Tissue Network (CHTN) or during surgeries.

The Affymetrix GeneChip Operating System (GCOS) was used to calculate expression summaries with a target intensity of 100 using Microarray Suite version 5.0 (MAS 5.0) (13). For completeness, the raw expression datasets were processed into expression estimates using the robust multiarray average method (14). The arrays were normalized using global scaling.

Differentially co-expressed genes (DCGs) and links (DCLs) were identified using the DCGL method (http://cran.r-project.org/web/packages/DCGL/index.html) (11,12), which was released as an R package that included two gene-filtering functions (an expression-based filter and a variance-based filter), three link-filtering functions (a systematic link filter, a percent link filter and a qlink filter) and five differential coexpression analysis functions (log ratio of connections [LRC], average specific connection [ASC], weighted gene coexpression network analysis [WGCNA], differential coexpression profile [DCp], and differential coexpression enrichment [DCe]). These functions generally use gene expression matrices (with genes in rows and samples in columns) as the major input, and the ultimate output is genes ranked by p-value. The p-values were adjusted for multiple comparisons using the false discovery rate (FDR) of Benjamini and Hochberg (15). An FDR-corrected p-values of 0.25 was used as the threshold value for screening the DCGs. The DCe has an additional output of classified DCLs.

KEGG pathway enrichment analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of online databases of genomes, enzymatic pathways, and biological chemicals (16). The PATHWAY database records networks of molecular interactions in the cells and variants that are specific to particular organisms

(http://www.genome.jp/kegg/). A total of 130 pathways including 2,287 genes were collected from KEGG (updated on 2011.06).

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (17), a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. DAVID was used to identify over-represented KEGG pathways based on the hypergeometric distribution. p-values <0.05 were considered statistically significant.

Analysis of transcription factors and their target genes

According to the central dogma, a variety of methods can lead to differential gene expression. Gene expression is regulated by TFs at the transcriptional level. Therefore, the roles of TFs were further analyzed using the acquired data. Information on the chromosome region to which human TFs bind was obtained from the UCSC database (http://genome.ucsc.edu). Subsequently, the chromosome annotations were downloaded from the NCBI database and analyzed to acquire the gene information of the above positions. Ultimately, a total of 210,000 regulatory pairs of TFs and their target genes were obtained. These pairs were then mapped to the above DCLs to determine the regulatory relationship between TFs and our DCGs.

■ RESULTS

Screening for differentially co-expressed genes

Using the DCGL method described above, the gene expression profiles of cells from discs with signs of degenerative disc disease and from herniated discs with different degrees were analyzed, and genes with FDR<0.25 were considered DCGs. A total of 539 DCGs and 113,866 significantly related pairs of DCLs were selected. These findings suggest that these selected DCGs and their connections may have crucial roles in the development of herniated discs and degenerative disc disease.

Analysis of biological pathways related to herniated discs and degenerative disc disease

A total of 539 DCGs were included in the KEGG pathway enrichment analysis, and significantly altered pathways were ranked according to their p-values. The results showed that these DCGs were mainly involved in the following pathways (Table 1): protein processing in endoplasmic reticulum, Alzheimer's disease, focal adhesion, glycosphingolipid biosynthesis-globo series, Huntington's disease, Parkinson's disease, oxidative phosphorylation, apoptosis, renal cell carcinoma, and selenoamino acid metabolism. Therefore, further analysis of these pathways may explain the pathogenesis of herniated discs and degenerative disc disease to some extent. The analysis of the apoptosis and ECM-receptor interaction pathways in particular may yield important insights because these pathways have been identified in previous studies.

Analysis of gene regulatory relationships

A total of 113,866 pairs of DCLs were matched with their already known TFs and their target genes, and 62 pairs of gene regulatory relationships (including 44 target genes) were found to be associated with herniated discs and

Table 1 - KEGG pathway enrichment analysis of differentially co-expressed genes.

ID	p-value	Count	Size	Term
4141	0.0011	11	167	Protein processing in endoplasmic reticulum
5010	0.0012	11	168	Alzheimer's disease
4510	0.0015	12	201	Focal adhesion
603	0.0032	3	14	Glycosphingolipid biosynthesis - globo series
5016	0.0074	10	184	Huntington's disease
5012	0.0087	8	132	Parkinson's disease
190	0.0094	8	134	Oxidative phosphorylation
4210	0.0130	6	88	Apoptosis
5211	0.0190	5	70	Renal cell carcinoma
450	0.0191	3	26	Selenoamino acid metabolism
5213	0.0275	4	52	Endometrial cancer
4512	0.0380	5	84	ECM-receptor interaction
5221	0.0433	4	60	Acute myeloid leukemia
5215	0.0469	5	89	Prostate cancer
511	0.0478	2	16	Other glycan degradation
5210	0.0480	4	62	Colorectal cancer

ID represents the identification number in the KEGG database; count represents the number of differentially co-expressed genes involved in this signaling pathway; size represents the number of total genes included in this pathway; and term represents the name of this KEGG pathway.

degenerative disc disease (Figure 1). This result indicated that the presence of these 62 pairs of gene regulatory relationships may lead to the differential expression of 44 target genes at the transcriptional level and ultimately result in the progression of disc herniation and of degenerative disc disease.

■ DISCUSSION

Based on our results described above, 539 DCGs were selected and shown to be enriched in several pathways in degenerative disc disease and herniated discs of different degrees. Among these pathways, apoptosis and ECM-receptor interaction (5) were demonstrated to be associated with these diseases. Both apoptosis and Fas/FasL expression have been reported to be present in human disc specimens (18,19). The Fas/FasL system is a surface ligand that leads to rapid cell death when a Fas-expressing cell interacts with a second cell carrying the FasL receptor. Fas expression is not observed in normal discs but can be observed following disc injury, leading to disc degeneration (20).

In addition, 113,866 pairs of DCLs were selected and matched with their known TFs and target genes. Finally, 62 pairs of regulatory relationships that led to the identification of differentially expressed genes associated with degenerative disc disease and herniated discs of different degrees were obtained. There were numerous biological regulatory relationships between TFs and target genes (Figure 1). When the TFs were DCGs, some of their target genes were DCGs and some were not, indicating that many other factors were synergistically involved in the regulation of these target genes. However, although the TFs were not DCGs, all of their target genes were DCGs. These TFs may regulate their target genes through changes in their spatial structures and configurations and changes in localization. For example, NF-êB TFs are kept inactive in the cytoplasm by IêB
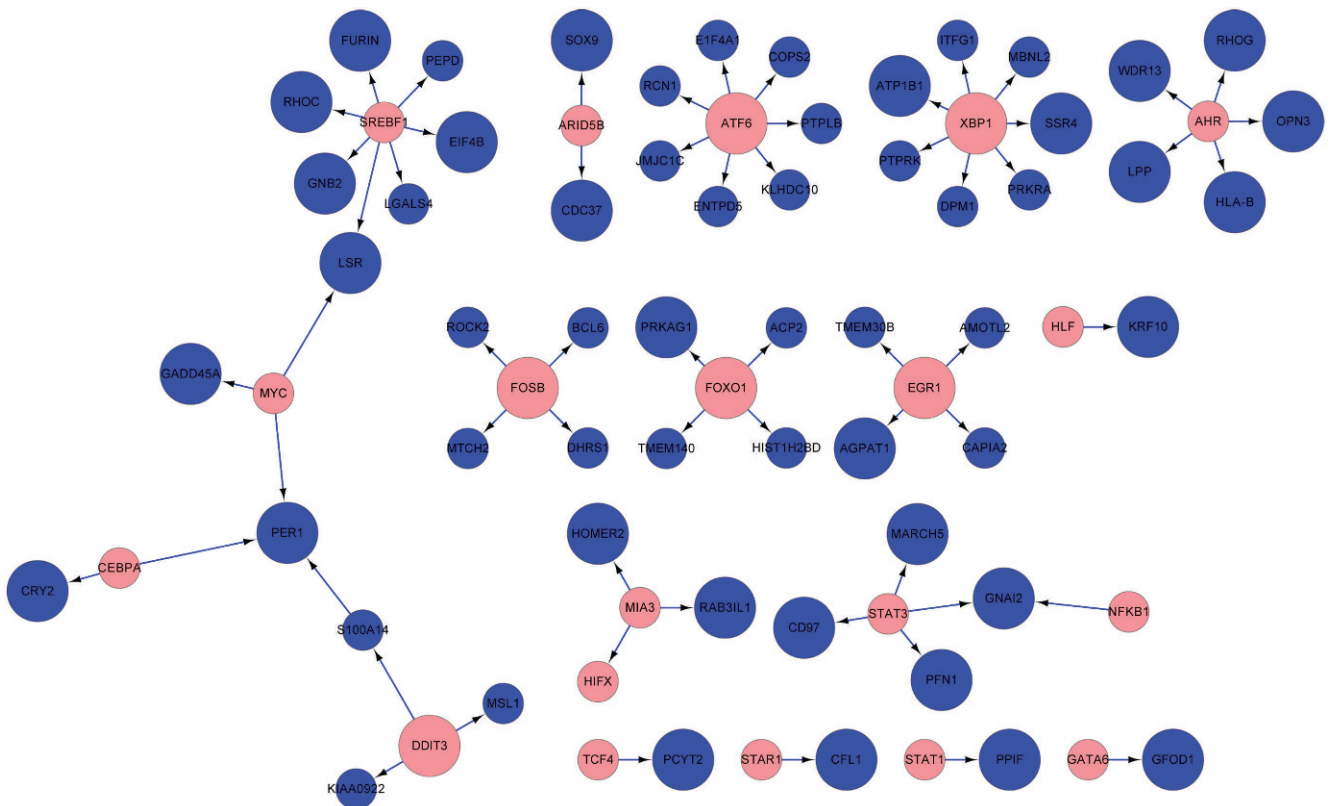

Figure 1 - Networks of transcription factors and their target genes included in set of 539 differentially co-expressed genes. Blue dots represent target genes, and red dots represent transcription factors. The larger dots represent differentially co-expressed genes, and the smaller dots represent genes that are not differentially co-expressed.

(inhibitors of NF-êB) proteins (21). The activation process is regulated by the IêB kinase (IKK) complex. IKK activation leads to the phosphorylation and degradation of IêB proteins, subsequent NF-êB nuclear translocation and the transcriptional initiation of NF-êB-dependent genes (22). Although the amount of TF (NF-êB) is not changed, the activated TF can regulate its target genes and lead to their differential expression.

In addition to these connections between TFs and target genes, many differentially expressed genes could still be regulated by TFs in indirect ways. For example, cathepsin K gene expression is significantly higher in more degenerated grade III to IV discs compared with healthier grade I to II discs ($p = 0.001$) (23). Cathepsin K, the protein encoded by CTSK, is a lysosomal cysteine protease involved in bone remodeling and resorption (24,25). Mice that are deficient in the cathepsin K gene have osteopetrosis of the long bones and vertebrae and abnormal joint morphology (26). Mutations in the human cathepsin K gene have been demonstrated to be associated with a rare skeletal dysplasia and pycnodysostosis (27,28). Cathepsin K over-expression could lead to Gaucher disease (29). In an analysis of CTSK, we found that CTSK could interact with the aryl hydrocarbon receptor (AHR) after interactions between eukaryotic translation initiation factor 2-alpha kinase 1 (EIF2AK) and ras homolog family member G (RHOG). The AHR protein contains several domains that are critical for its function and is classified as a member of the basic helix-loop-helix/Per-Arnt-Sim (bHLH/PAS) family of TFs (30,31). Therefore, although there is no direct relationship between CTSK and TFs, as shown in in Figure 1, CTSK can be regulated through interactions with other molecules (AHR->RHOG---EIF2AK1---CTSK). It can be inferred that CTSK can be regulated by several types of upstream genes, resulting in differential expression.

In conclusion, DCGs and the corresponding pathways, such as apoptosis and ECM-receptor interaction, were demonstrated to be involved in degenerative disc disease and herniated discs of different grades. Additionally, critical regulatory relationships between TFs and target genes, including one gene and several upstream regulators, were also identified in our study. Further experiments will be necessary to confirm these conclusions.

■ AUTHOR CONTRIBUTIONS

■ REFERENCES

1. Andersson GB. Epidemiological features of chronic low-back pain. Lancet. 1999;354(9178):581-5, http://dx.doi.org/10.1016/S0140-6736(99)01312-4.

2. Somanath PR, Razorenova OV, Chen J, Byzova TV. Akt1 in endothelial cell and angiogenesis. Cell Cycle. 2006;5(5):512-8, http://dx.doi.org/10.4161/cc.5.5.2538.

3. Schwarzer AC, Aprill CN, Derby R, Fortin J, Kine G, Bogduk N. The relative contributions of the disc and zygapophyseal joint in chronic low back pain. Spine. 1994;19(7):801-6, http://dx.doi.org/10.1097/00007632-199404000-00013.

4. Adams MA, Roughley PJ. What is intervertebral disc degeneration, and what causes it? Spine. 2006;31(18):2151-61, http://dx.doi.org/10.1097/01.brs.0000231761.73859.2c.

5. Urban JP, Roberts S. Degeneration of the intervertebral disc. Arthritis Res Ther. 2003;5(3):120-30, http://dx.doi.org/10.1186/ar629.

6. Colombini A, Lombardi G, Corsi MM, Banfi G. Pathophysiology of the human intervertebral disc. Int J Biochem Cell Biol. 2008;40(5):837-42, http://dx.doi.org/10.1016/j.biocel.2007.12.011.

7. Johnson WE, Eisenstein SM, Roberts S. Cell cluster formation in degenerate lumbar intervertebral discs is associated with increased disc cell proliferation. Connective Tissue Res. 2001;42(3):197-207, http://dx.doi.org/10.3109/03008200109005650.

8. Battie MC, Videman T, Levalahti E, Gill K, Kaprio J. Genetic and environmental effects on disc degeneration by phenotype and spinal level: a multivariate twin study. Spine. 2008;33(25):2801-8, http://dx.doi.org/10.1097/BRS.0b013e31818043b7.

9. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. Genome Res. 2004;14(6):1085-94, http://dx.doi.org/10.1101/gr.1910904.

10. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science. 2003;302(5643):249-55, http://dx.doi.org/10.1126/science.1087447.

11. Yu H, Liu BH, Ye ZQ, Li C, Li YX, Li YY. Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. BMC Bioinformatics. 2011;12:315, http://dx.doi.org/10.1186/1471-2105-12-315.

12. Liu BH, Yu H, Tu K, Li C, Li YX, Li YY. DCGL: an R package for identifying differentially coexpressed genes and links from gene expression microarray data. Bioinformatics. 2010;26(20):2637-8, http://dx.doi.org/10.1093/bioinformatics/btq471.

13. Hubbell E, Liu WM, Mei R. Robust estimators for expression analysis. Bioinformatics. 2002;18(12):1585-92, http://dx.doi.org/10.1093/bioinformatics/18.12.1585.

14. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249-64, http://dx.doi.org/10.1093/biostatistics/4.2.249.

15. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Statis Society Series B (Methodological). 1995:289-300.

16. Kanehisa M. The KEGG database. Novartis Found Symp. 2002;247:91-101; discussion -3, 19-28, 244-52, http://dx.doi.org/10.1002/0470857897.ch8.

17. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.

18. Park JB, Chang H, Kim KW. Expression of Fas ligand and apoptosis of disc cells in herniated lumbar disc tissue. Spine. 2001;26(6):618-21, http://dx.doi.org/10.1097/00007632-200103150-00011.

19. Park JB, Kim KW, Han CW, Chang H. Expression of Fas receptor on disc cells in herniated lumbar disc tissue. Spine. 2001;26(2):142-6, http://dx.doi.org/10.1097/00007632-200101150-00006.

20. Anderson DG, Izzo MW, Hall DJ, Vaccaro AR, Hilibrand A, Arnold W, et al. Comparative gene expression profiling of normal and degenerative discs: analysis of a rabbit annular laceration model. Spine. 2002;27(12):1291-6, http://dx.doi.org/10.1097/00007632-200206150-00009.

21. Novotny NM, Markel TA, Crisostomo PR, Meldrum DR. Differential IL-6 and VEGF secretion in adult and neonatal mesenchymal stem cells: role of NFkB. Cytokine. 2008;43(2):215-9, http://dx.doi.org/10.1016/j.cyto.2008.05.015.

22. Karin M, Ben-Neriah Y. Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity. Annual Rev Immunol. 2000;18:621-63, http://dx.doi.org/10.1146/annurev.immunol.18.1.621.

23. Gruber HE, Ingram JA, Hoelscher GL, Zinchenko N, Norton HJ, Hanley EN, Jr. Constitutive expression of cathepsin K in the human inter-vertebral disc: new insight into disc extracellular matrix remodeling via cathepsin K and receptor activator of nuclear factor-kappaB ligand. Arthritis Res Ther. 2011;13(4):R140, http://dx.doi.org/10.1186/ar3454.

24. Inaoka T, Bilbe G, Ishibashi O, Tezuka K, Kumegawa M, Kokubo T. Molecular cloning of human cDNA for cathepsin K: novel cysteine proteinase predominantly expressed in bone. Biochem Biophy Res Commun. 1995;206(1):89-96, http://dx.doi.org/10.1006/bbrc.1995.1013.

25. Salminen-Mankonen HJ, Morko J, Vuorio E. Role of cathepsin K in normal joints and in the development of arthritis. Current Drug Targets. 2007;8(2):315-23, http://dx.doi.org/10.2174/138945007779940188.

26. Gowen M, Lazner F, Dodds R, Kapadia R, Feild J, Tavaria M, et al. Cathepsin K knockout mice develop osteopetrosis due to a deficit in matrix degradation but not demineralization. J Bone Mineral Res. 1999;14(10):1654-63, http://dx.doi.org/10.1359/jbmr.1999.14.10.1654.

27. Gelb BD, Shi GP, Chapman HA, Desnick RJ. Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. Science. 1996;273(5279):1236-8, http://dx.doi.org/10.1126/science.273.5279.1236.

28. Johnson MR, Polymeropoulos MH, Vos HL, Ortiz de Luna RI, Francomano CA. A nonsense mutation in the cathepsin K gene observed in a family with pycnodysostosis. Genome Res. 1996;6(11):1050-5, http://dx.doi.org/10.1101/gr.6.11.1050.

29. Moran MT, Schofield JP, Hayman AR, Shi GP, Young E, Cox TM. Pathologic gene expression in Gaucher disease: up-regulation of cysteine proteinases including osteoclastic cathepsin K. Blood. 2000;96(5):1969-78.

30. Burbach KM, Poland A, Bradfield CA. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. Proc Natl Acad Sci U S A. 1992;89(17):8185-9, http://dx.doi.org/10.1073/pnas.89.17.8185.

31. Fukunaga BN, Probst MR, Reisz-Porszasz S, Hankinson O. Identification of functional domains of the aryl hydrocarbon receptor. J Biol Chem. 1995;270(49):29270-8.