Analysis of intervertebral disc-related genes

D.J. Wu, K. Chen, X.Z. Wei, H.J. Ni, S.Z. Yu, X.D. Zhu and M. Li

Department of Orthopedic,
Changhai Hospital of the Second Military Medical University,
Shanghai, China

Corresponding author: M. Li
E-mail: limingllmm@hotmail.com

ABSTRACT. Intervertebral disk disease is a common clinical disorder manifested by pain, ataxia, paresis, motor paralysis, and sensorimotor paralysis. The clinical features, diagnosis, and treatment of cervical and thoracolumbar disk disease have been unclear until now. In this study, some differentially expressed genes were identified, and a network was constructed based on these genes. Through the statistical analysis of nodes and the contrast of 2 more connectivity nodes, it was found that the nodes in the network are in an important position and play key roles. Several of these genes, including MAP2K6, MAP2K3, and MAPK14, belong to the MAP kinase family, and several genes, including RHOBTB2, RHOQ, and RHOH, belong to the RHO family. Therefore, we hypothesize that the development of intervertebral disk disease is related to MAP and RHO family proteins.

Key words: Intervertebral disk disease; MAP kinase family;
Rho family

INTRODUCTION

Intervertebral disc disorder is a condition that involves deterioration, herniation, or other dysfunctions of the intervertebral disc (Danielsen et al., 2011). The human spinal column is composed of 33 vertebrae that are in stacked bone segments: 7 cervical, 12 thoracic, 5 lumbar, 5 sacral (fused), and 2 to 4 coccygeal (fused) (Gambardella, 2008). The intervertebral discs are structures that lie between the individual vertebrae and act as a cushion between the discs. When the nucleus pulposus starts to dehydrate, its ability to transfer and distribute loads between the vertebrae is being restricted, which will cause chronic pain (Savolainen et al., 1998).

The clinical features are unclear and the diagnosis and treatment of cervical and thoracolumbar disk disease are challenging. Clinical signs are determined by neuroanatomic localization and the severity of the spinal cord injury (Taveras and Wood, 1976; De Lahunta and Glass, 2008). Myelography is most commonly used in clinical diagnosis; however, its accuracy is not determined (Bolender et al., 1985). Thus, recent studies have provided a new understanding of the diagnosis and treatment of intervertebral disc disorder.

Based on intervertebral disc-related mRNA expression profile data, we obtained differentially expressed mRNAs between the third and fourth stages of the disease compared with the second stage of the intervertebral disc disease. In addition, we built two interaction networks at the center of the target genes. Finally, we used GO analysis to perform statistical analysis of the network nodes and to find the genes which may play important roles at the different stages of intervertebral disc disease.

MATERIAL AND METHODS

Expression profiles of intervertebral disc disease-related genes

To recognize differentially expressed genes of intervertebral disc disease in 3 different stages from the molecular level and the biological network level, we downloaded the GSE15277 data set from the Gene Expression Omnibus (GEO: http://www.ncbi.nlm.nih.gov/geo/) database. The experiment tested the expression of intervertebral disc in different stages, including 15 samples (5 second-stage samples, 7 third-stage samples, and 3 fourth-stage samples).

Differential expression analysis

First, we converted the original data into recognizable files and then filled in the lacking data. Next, all data were normalized using log (Ration) with the affy package of the *R* language (Ihaka and Gentleman, 1996). The mRNA expression profiles were used to perform differential expressional analysis, using the *t*-test to recognize differentially expressed genes between the third and fourth stages of the disease compared with the second stage of the disease. To verify that the P values were closely related to the fold-change, we used the volcano plot to reflect the P values directly. P values were corrected by the BH (Beyer-Hardwick) method, and a threshold P value of 0.05 was defined (Smyth, 2004).

Interaction network mapping

We numbered the downloaded probes with the ID converter and used the STRING

(Search Tool for the Retrieval of Interacting Genes/Proteins) database to predict the interaction of objects of a protein encoded by a target gene. The results of the 2 network node analyses suggested the role that the target genes may play in the network built from sequence characteristics, species, and structure.

GO analysis

We annotated the nodes of the interactional network map using GO database (Ashburner et al., 2000; Wang et al., 2012). Finally, we identified the genes that may play key roles at the different stages of intervertebral disc disease.

RESULTS

Differential expression analysis

Due to various reasons, such as background and probe design challenges, there is a great difference in quality of the original data. Thus, we had to perform normalization of the data.

The *t*-test and BH multiple test correction were used to conduct differential expression analysis of the genes, and the threshold was set at 0.5. We selected $P < 0.05$ as the subsequent significance level for identifying differentially expressed genes (Table 1 and Figure 1).

Table 1. Differentially expressed genes.

ID_REF	2v3		
	t-test	adj	logFC
g10835026_3p_a_at	1.52E-06	0.030373	0.898943
g6642743_3p_at	1.98E-06	0.030373	0.897715
Hs.21236.0.A1_3p_at	3.58E-06	0.043933	-0.48729
Hs.296702.0.S1_3p_at	9.03E-07	0.027704	0.821719
Hs.55879.2.A1_3p_s_at	4.32E-07	0.026507	-0.51696
	2v4		
g4426566_3p_at	4.41E-06	0.031723	1.000226
g5453631_3p_at	9.61E-06	0.044704	0.645519
g5453813_3p_at	7.07E-07	0.02301	0.751899
g7656941_3p_at	5.16E-06	0.031723	0.589652
g8923438_3p_s_at	5.17E-06	0.031723	0.841551
g9845486_3p_a_at	6.98E-06	0.038935	0.549135
Hs.10487.0.S1_3p_at	1.35E-06	0.02301	-1.00635
Hs.118021.3.A1_3p_a_at	3.07E-06	0.031723	0.661216
Hs.132809.0.A1_3p_at	1.26E-06	0.02301	-0.9397
Hs.291891.0.A1_3p_at	1.50E-06	0.02301	-0.97467
Hs.324178.0.S2_3p_at	4.12E-06	0.031723	0.844295
Hs.98267.0.A1_3p_at	3.91E-06	0.031723	0.649635
Hs.98485.0.S1_3p_a_at	1.02E-05	0.044704	0.574258
Hs2.62113.2.S1_3p_s_at	9.43E-06	0.044704	0.658343

In order to verify that the P value obtained by different ways and value of the fold-change (FC) were logical, we used the volcano plot (Figure 1). The smaller the P value, the greater the fold-change.

Conversion of the probe number to the corresponding gene name

The raw data only provided the gene number; therefore, the probe numbers were converted to the corresponding gene names (Table 2).
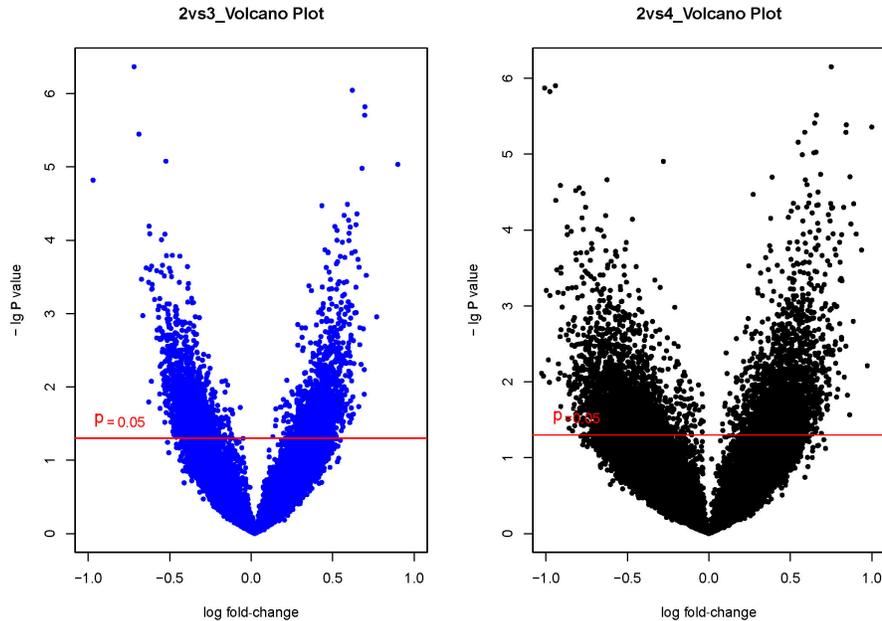
Figure 1. Volcano plot verification.

Table 2. Conversion of the probe number and corresponding gene name.

Gene name	Ensembl gene	UniGene	EntrezGene	Affymetrix	GenBank accession	Ensembl* Chr
MAP2K6	ENSG00000108984		6846	[U133_X3P] g10835026_3p_a_at		17
	ENSG00000177320			[U133_X3P] g6642743_3p_at		2
				Hs.21236.0.A1_3p_at		
				Hs.296702.0.S1_3p_at		
ABCC10	ENSG00000124574	Hs.55879	52	[U133_X3P]	NM_033450	6
				Hs.55879.2.A1_3p_s_at		
CLIC1	ENSG00000213719	Hs.414565	1192	[U133_X3P] g4426566_3p_at	BC064527 BQ051912	6
					BM558439	
SPAG5	ENSG00000076382	Hs.514033	10615	[U133_X3P] g5453631_3p_at	AF345347 AF399910	17
					NM_006461	
DDR2	ENSG00000162733	Hs.593833	2731	[U133_X3P] g5453813_3p_at	BC052998	1
					NM_001014796	
	ENSG00000100211	Hs.696291	25776	[U133_X3P] g7656941_3p_at		22
	ENSG00000114544		54946	[U133_X3P] g8923438_3p_s_at		3
	ENSG00000072518		2011	[U133_X3P] g9845486_3p_a_at		11
CCDC16	ENSG00000198783	Hs.413678 Hs.691125	91603	[U133_X3P] Hs.10487.0.S1_3p_at	NM_052857 CR617791	17
ABR	ENSG00000159842	Hs.159306	29	[U133_X3P]	AK124547 NM_021962	17
				Hs.118021.3.A1_3p_a_at	U01147 NM_001092	
				Hs.132809.0.A1_3p_at		
				Hs.291891.0.A1_3p_at		
MAPKAP1	ENSG00000119487	Hs.495138 Hs.597362	79109	[U133_X3P]	NM_001006617	9
		Hs.629072		Hs.324178.0.S2_3p_at	AY524429	
					NM_024117	
				Hs.98267.0.A1_3p_at		

Construction and analysis of network

First, we summarized all of the nodes of each network and selected the nodes that

connectivity warranted further analysis (Figure 2). Through statistical analysis of the nodes, and by comparing to 2 additional connectivity nodes, we found unique network nodes and high-value nodes, which indicated that these nodes were located in an important position and played key roles in the network (Table 3).
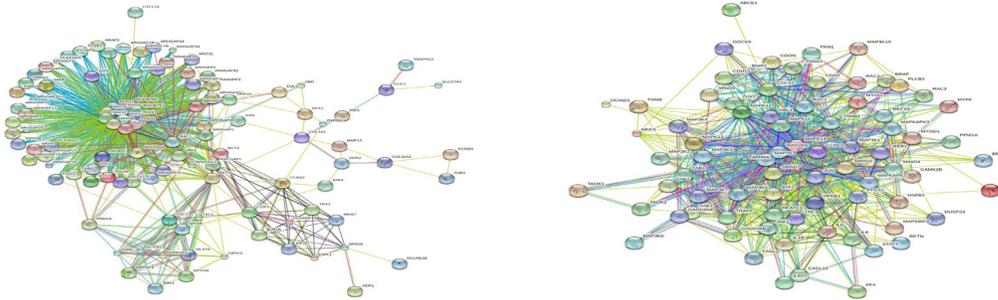
Figure 2. Network of differentially expressed genes.

Table 3. Top ten genes in the network.

Stage 2 *vs* Stage 3		Stage 3 *vs* Stage 4	
Gene	Degree	Gene	Degree
MAP2K6	96	RAC2	70
MAP2K3	63	RHOB	68
MAPK14	63	RHOBTB2	68
JUN	58	RHOQ	68
TNF	46	RHOG	68
NFKB1	44	RHOBTB1	67
FAM48A	44	RHOH	67
MAPK11	41	RHOV	66
MAP3K1	39	RHOD	66
CDC42	37	RHOU	66

In these genes, MAP2K6, MAP2K3, MAPK14, etc. belong to the MAP kinase family, as well as RHOBTB2, RHOQ, RHOH, etc. belong to the RHO family. So we considered that the development of intervertebral disk disease is related to MAP and RHO families.

DISCUSSION

Intervertebral disk disease is a common clinical disorder manifested by pain, ataxia, paresis, motor paralysis, and sensorimotor paralysis (Tokuhashi et al., 2001). The literature by Hoerlein, Olsson, Hansen, Funquist, and many others, formed the foundation of our current medical and surgical therapies for intervertebral disk disease protrusion (Shores, 1981). Numerous manuscripts are related to surgical technique, radiologic diagnosis, pathology, pathophysiology, and incidence of intervertebral disk disease (Bagley, 1999). Furthermore, the proteoglycan and interstitial fluid content is known to decrease with age in the anulus fibrosus, decreasing the ability to absorb shock.

In our study, we obtained some differentially expressed genes between 2 *vs* 3 and 2 *vs* 4, including g10835026_3p_a_at, g6642743_3p_at, g4426566_3p_at, and g5453631_3p_at. Through statistical analysis of the nodes in the networks as well as comparison to 2 connectivity nodes, we found nodes with important positions that may play key roles.

Indications for surgical management are pain or paresis unresponsive to medical therapy and recurrent or progressive signs of thoracolumbar intervertebral disk disease (Gambardella, 2008). The major postoperative considerations are based on the surgeon's experience, ability, and preference (McPhee et al., 2010). Most of these concerns are valid also for medically managed intervertebral disk disease patients.

Biochemically, intervertebral disk disease consists of proteoglycans, glycoproteins, and both collagenous and noncollagenous proteins (Diwan et al., 2000). The therapeutic measures of choice for intervertebral disk disease have long been debated (Sussman, 1975). A plethora of genes has been identified that are related to intervertebral disc disease, including genes related to intervertebral disc stability, inflammation, and signaling pathways. In our analysis, we used expression profile data to analyze differentially expressed genes at different stages of intervertebral disk disease and found that many of the key genes (nodes on the network) are relevant to signaling pathways.

Medical therapy and surgical therapy seek to alleviate the pain or neurologic deficits associated with intervertebral disk disease. Degeneration of intervertebral disk disease occurs with aging and results in significant changes in biochemical characteristics. The progress of intervertebral disc disease is regulated by a large combination of genes.

The progress of intervertebral disk disease is regulated by multiple genes. Studied and unknown genes related to intervertebral disc disease are likely involved in the occurrence and development of this disease. Therefore, it is necessary to develop long-term, large sample case-control studies of different groups to aid the study of intervertebral disc disease genetic mechanisms. These studies would allow personalized prevention and treatment of intervertebral disk disease.

RHOBTB2, RHOQ, and RHOH are members of the mitogen-activated protein (MAP) kinase family, also known as the extracellular signal-regulated kinase family. The MAP family is involved in many signaling cascades. MAP family members are phosphorylated when they respond to inflammatory cytokines or environment stress. The p38 protein, an important part of the MAP kinase-mediated pathway, participates in many cellular processes such as cell cycle arrest, transcriptional activation, and apoptosis (Johnson and Lapadat, 2002; Olek et al., 2003; Diao et al., 2010; Danielsen et al., 2011). The pathologic changes associated with acute spinal cord compression are related to dynamic force of compression.

In signaling pathways, the RHO family of proteins plays the role of a molecular switch by influencing the cytoskeleton. The RHOBTB2 proteins can participate in signal transduction pathways and regulate cell growth by affecting cell apoptosis and the cell cycle.

REFERENCES

Ashburner M, Ball CA, Blake JA, Botstein D, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25-29.

Bagley RS (1999). Controversies in Intervertebral Disk Disease. In: ACVS Veterinary Symposium: Large Animal, San Francisco.

Bolender NF, Schönström NS and Spengler DM (1985). Role of computed tomography and myelography in the diagnosis of central spinal stenosis. *J. Bone Joint Surg. Am.* 67: 240-246.

Danielsen JM, Sylvestersen KB, Bekker-Jensen S, Szklarczyk D, et al. (2011). Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell Proteomics* 10: M110.

De Lahunta A and Glass E (2008). Veterinary Neuroanatomy and Clinical Neurology. WB Saunders Company, Philadelphia.

Diao Y, Liu W, Wong CC, Wang X, et al. (2010). Oxidation-induced intramolecular disulfide bond inactivates mitogen-activated protein kinase kinase 6 by inhibiting ATP binding. *Proc. Natl. Acad. Sci. U. S. A.* 107: 20974-20979.

Diwan AD, Parvataneni HK, Khan SN, Sandhu HS, et al. (2000). Current concepts in intervertebral disc restoration. *Orthop. Clin. North Am.* 31: 453-464.

Gambardella PC (2008). Dorsal decompressive laminectomy for treatment of thoracolumbar disc disease in dogs: a retrospective study of 98 cases. *Vet. Surg.* 9: 24-26.

Ihaka R and Gentleman R (1996). R: A language for data analysis and graphics. *J. Comp. Graphical Stat.* 5: 299-314.

Johnson GL and Lapadat R (2002). Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 298: 1911-1912.

McPhee SJ, Papadakis MA and Rabow MW (2010). Current Medical Diagnosis & Treatment 2010. McGraw-Hill Medical, New York.

Olek A, Piepenbrock C and Berlin K (2003). Method for Detecting Methylation States for a Toxicological Diagnostic, Google Patents.

Savolainen S, Rinne J and Hernesniemi J (1998). A prospective randomized study of anterior single-level cervical disc operations with long-term follow-up: surgical fusion is unnecessary. *Neurosurgery* 43: 51-55.

Shores A (1981). Intervertebral disk disease. *Veterinarian* 3: 639-647.

Smyth GK (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3: Article 3.

Sussman BJ (1975). Inadequacies and hazards of chymopapain injections as treatment for intervertebral disc disease. *J. Neurosurg.* 42: 389-396.

Taveras JM and Wood EH (1976). Diagnostic Neuroradiology. Williams & Wilkins Company, Philadelphia.

Tokuhashi Y, Matsuzaki H, Uematsu Y and Oda H (2001). Symptoms of thoracolumbar junction disc herniation. *Spine* 26: E512-E518.

Wang J, Zhou X, Zhu J and Gu Y (2012). GO-function: deriving biologically relevant functions from statistically significant functions. *Brief. Bioinform.* 13: 216-227.