GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genes associated with disc degeneration identified using microarray gene expression profiling and bioinformatics analysis

Y. Chen*, K. Chen*, M. Li, C. Li, H. Ma, Y.S. Bai, X.D. Zhu and Q. Fu

Changhai Hospital, Second Military Medical University, Shanghai, China

*These authors contributed equally to this study.
Corresponding author: M. Li
E-mail: minglilimingmmll@hotmail.com

Genet. Mol. Res. 12 (2): 1431-1439 (2013)
Received July 30, 2012
Accepted December 20, 2012
Published April 26, 2013
DOI http://dx.doi.org/10.4238/2013.April.26.5

ABSTRACT. Disc degeneration is strongly associated with back or neck pain, sciatica, and disc herniation or prolapse. It places an enormous economic burden on society and can greatly affect quality of life. Alternative treatment approaches, such as genetic therapies, are urgently needed to slow or reverse the disc degeneration process. We downloaded gene expression data from Gene Expression Omnibus during various stages of disc degeneration and identified differentially expressed genes (DEGs) as well as dysfunctional pathways through comparisons with controls. We identified 2 significant DEGs between grade II and III discs and 8 significant DEGs between grade II and IV discs. By constructing an interactive network of the DEGs, we found that mitogen-activated protein family genes and Ras homologous (Rho) family genes - in particular, *MAP2K6* and *RHOBTB2* - may play important roles in the progression of degeneration of grade III and IV discs, respectively. *MAP2K6* and *RHOBTB2* may be specific therapeutic molecular targets in the treatment of disc degeneration. However, further experiments are needed to confirm this result.

Key words: Disc degeneration; Differentially expressed gene; Dysfunctional pathway

INTRODUCTION

Disc degeneration in the human spine is a complex phenomenon characterized by biochemical change in the nucleus pulposus and inner annulus and the formation of clefts and fissures (Fraser et al., 1993). It alters disc height and the mechanics of the rest of the spinal column, possibly adversely affecting the behavior of other spinal structures such as muscles and ligaments (Urban and Roberts, 2003). Disc degeneration is strongly associated with back or neck pain, sciatica, and disc herniation or prolapse. It places an enormous economic burden on society and can greatly affect quality of life (Luoma et al., 2000).

Disc degeneration is a multifactorial process influenced by mechanical pressures, aging, genetic inheritance, and alterations in the cellular biology of intervertebral discs, among other factors (Freemont et al., 1997; Hadjipavlou, et al., 2008). Disc degeneration might originate from the injury and subsequent repair of the annulus fibrosus (Peng et al., 2006). Growth factors, such as transforming growth factor, interleukin, insulin-like growth factor, and basic fibroblast growth factor, through each of their receptor signal transduction pathways, promote cellular proliferation and collagen synthesis of matrix cells and play a key role in disc degeneration (Thompson et al., 1991; Nagano et al., 1995; Doita et al., 1996). By contrast, cytokines, such as matrix metalloproteinases (MMPs) and a disintegrin-like and metalloproteinase with thrombospondin motifs (ADAMTS), inhibit matrix synthesis and stimulate the production of degradative enzymes that break down the extracellular matrix of the intervertebral disc (IVD) (Le Maitre et al., 2007a). Increased MMP and ADAMTS enzyme activity - especially that of MMP7, MMP13, ADAMTS4, and ADAMTS5 - is a characteristic of disc degeneration (Le Maitre et al., 2004, 2006, 2007b). In addition, poor transport of nutrients into the disc is also a cause of degeneration (Nachemson et al., 1970).

In this study, we aimed to identify the differentially expressed genes (DEGs) in the progression of disc degeneration and gain insights into the molecular mechanisms of disc degeneration. Understanding these mechanisms can aid in selecting an appropriate treatment strategy and developing new treatments for degenerated discs.

MATERIAL AND METHODS

DNA microarray data

We extracted the gene expression profiles from the study of Gruber et al. (2009), which are available in the Gene Expression Omnibus database (ID: GSE15227). The study had been carried out to determine the localization patterns of aspirin expression in human discs. Disc degeneration was scored using the Thompson scoring system, which scores disc degeneration over the spectrum from healthy (grade I) to advanced degeneration (grade V, the most advanced stage of degeneration) (Thompson et al., 1990). Patient specimens were derived from surgical disc procedures performed on individuals with herniated discs and degenerative disc disease. A total of 15 chips were available, including 5 grade II discs, 7 grade III discs, and 3 grade IV discs.

Data preprocessing

The probe-level data in CEL files were converted into expression measures, and back-

ground correction and quartile data normalization were performed using the robust multi-array average (Irizarry et al., 2003) algorithm with the defaulted parameters in the Raffy package (Gautier et al., 2004; Team, 2011).

DEG analysis

The *t*-test was used to identify genes that were significantly differentially expressed between healthy discs (grade II) and unhealthy discs (grades III, IV). Probe sets were mapped to National Center for Biotechnology Information Entrez genes using Gene ID converter (Alibés et al., 2007). If multiple probe sets corresponded to the same gene, the expression values of those probe sets were averaged. The P value was adjusted using the Benjamini and Hochberg (Benjamini, 1995) method based on the multtest package (van der Laan et al., 2004); a cut-off of 0.05 was used.

Network analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) database was used for both experimental and predicted interaction information (Szklarczyk et al., 2011). Version 9.0 of STRING covers more than 1100 completely sequenced organisms, and all associations are provided with a probabilistic confidence score, which is derived by separately benchmarking groups of associations against the manually curated functional classification scheme of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Each score represents a rough estimate of how well a given association describes a functional linkage between 2 proteins that is at least as specific as that between an average pair of proteins annotated on the same "map" or "pathway" in KEGG. We used the STRING database to annotate functional interactions between DEGs and other genes by calculating their confidence score. The networks were built using Cytoscape (Shannon et al., 2003).

Pathway-enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery now provides a comprehensive set of functional annotation tools for understanding the biological meaning behind large list of genes (Huang et al., 2009). For functional annotation of genes in the interaction network, we identified the overrepresented KEGG categories in pathways. A count number larger than 2 and a false-discovery rate of <0.05 were chosen as cut-off criteria.

RESULTS

DEG analysis

We obtained the microarray dataset GSE15227 from the public Gene Expression Omnibus database. The *t*-test was used to identify genes specifically differentially expressed between healthy and unhealthy discs with multiple testing corrections. At an adjusted P value of 0.05, 10 genes showed a significant differential expression (Table 1), including 2 genes between grades II and III and 8 genes between grades II and IV.

Table 1. Differentially expressed genes (P value <0.05, adjusted P value <0.05).

	Gene	P	Adjusted P
Grade II *vs* Grade III	*MAP2K6*	1.52E-06	0.03037
	ABCC10	4.32E-07	0.02651
Grade II *vs* Grade IV	*CLIC1*	4.41E-06	0.03172
	SPAG5	9.61E-06	0.0447
	DDR2	7.07E-07	0.02301
	ZNF830	1.35E-06	0.02301
	ABR	3.07E-06	0.03172
	MAPKAP1	4.12E-06	0.03172
	GJB3	1.02E-05	0.0447
	RHOBTB2	9.43E-06	0.0447

Construction of a DEG interaction network

We mapped the DEGs to the STRING database and screened significant interactions with scores larger than 0.9. By integrating these relationships, we constructed interaction networks between DEGs and their interactive genes in grade III discs (Figure 1) and grade IV discs (Figure 2). We found that mitogen-activated protein (MAP) kinase kinase 6 (*MAP2K6*) is the hub node in Figure 1, and Ras homologous-related BTB domain containing 2 (*RHOBTB2*) is the hub node in Figure 2.

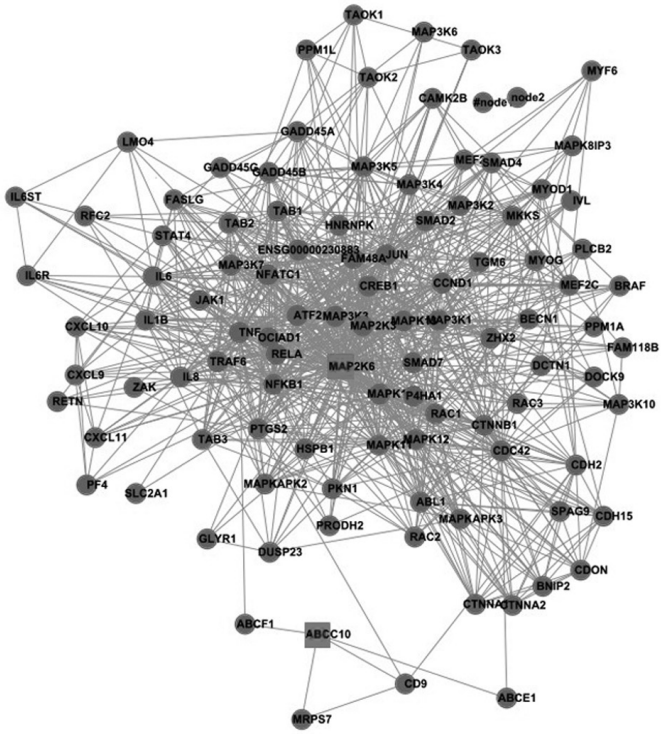

Figure 1. Interaction network constructed in grade III discs. The squares are differentially expressed genes, and the nodes are their interactive genes.

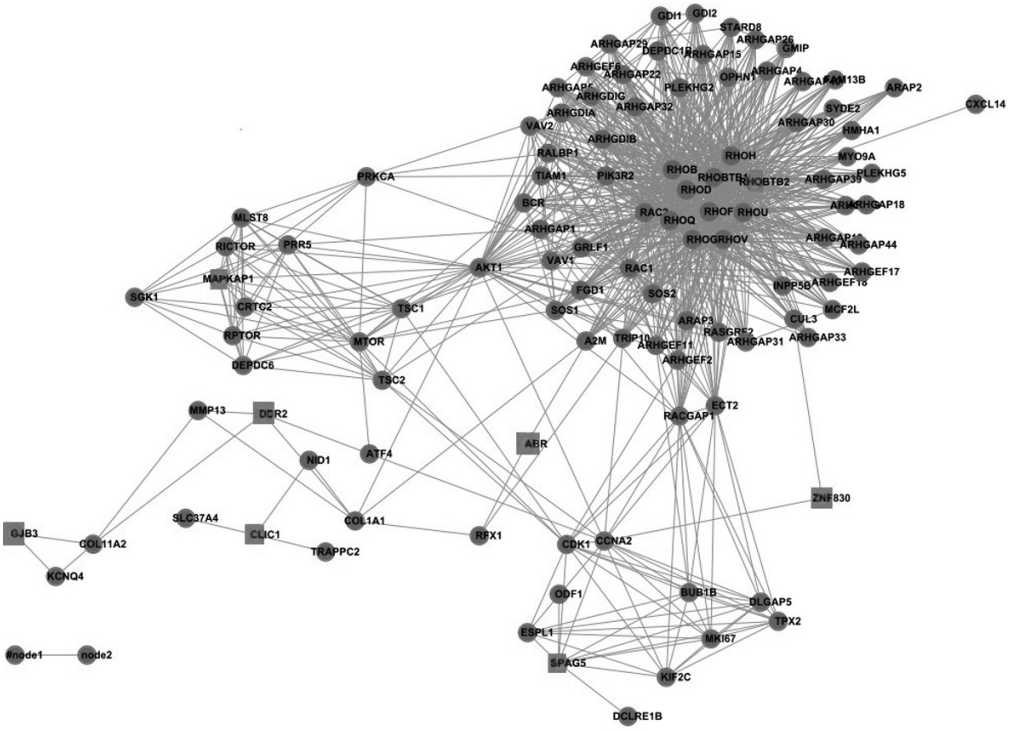

Figure 2. Interaction network constructed in grade IV discs. The squares are differentially expressed genes, and the nodes are their interactive genes.

We then calculated the average degree of each node in the networks. The average degree is the average number of edges connecting all the nodes in the network. A higher average degree value indicates a better connected network that is likely more robust. The average degrees of *MAP2K6* and *RHOBTB2* were 96 and 68, respectively. In addition, the average degree of genes in the MAP family [such as *MAP2K3* and MAP kinase (MAPK)14] and Ras homologous (Rho) family (such as *RHOB*, *RHOQ*, *RHOG*, and *RHOBTB1*) were higher than those of other genes in Figures 1 and 2.

Pathway-enrichment analysis of the interaction network

To gain further insights into the function of genes in our interaction network, we used the online biological classification tool Database for Annotation, Visualization and Integrated Discovery and observed significant enrichment of these genes in multiple KEGG terms (Tables 2 and 3). The most significant enrichment pathway in grade III discs was the MAPK signaling pathway. Other significant pathways included the Toll- and NOD-like receptor signaling pathway. The most significant enrichment pathway in grade IV discs was focal adhesion. The other significant pathways included the mTOR signaling and Fc epsilon RI signaling pathways.

Table 2. Enriched Kyoto Encyclopedia of Genes and Genome pathways in grade III discs.

Term	Description	P	FDR
hsa04010	MAPK signaling pathway	2.02E-32	2.19E-29
hsa04620	Toll-like receptor signaling pathway	4.16E-18	4.52E-15
hsa04621	NOD-like receptor signaling pathway	1.09E-13	1.18E-10
hsa04722	Neurotrophin signaling pathway	1.14E-12	1.24E-09
hsa04370	VEGF signaling pathway	5.35E-10	5.80E-07
hsa04912	GnRH signaling pathway	1.07E-09	1.16E-06
hsa04622	RIG-I-like receptor signaling pathway	4.28E-09	4.64E-06
hsa05200	Pathways in cancer	4.55E-09	4.94E-06
hsa05212	Pancreatic cancer	6.80E-08	7.38E-05
hsa05120	Epithelial cell signaling in *Helicobacter pylori* infection	4.96E-07	5.39E-04
hsa05014	Amyotrophic lateral sclerosis (ALS)	7.97E-07	8.65E-04
hsa04520	Adherens junction	1.46E-06	0.001586
hsa04664	Fc epsilon RI signaling pathway	1.63E-06	0.001772
hsa04660	T cell receptor signaling pathway	3.21E-06	0.003488
hsa04310	Wnt signaling pathway	1.05E-05	0.011383
hsa05210	Colorectal cancer	2.72E-05	0.029512

FDR = false-discovery rate.

Table 3. Enriched Kyoto Encyclopedia of Genes and Genomes pathways in grade IV discs.

Term	Description	P value	FDR
hsa04510	Focal adhesion	9.56E-08	9.62E-05
hsa04150	mTOR signaling pathway	2.60E-07	2.61E-04
hsa04664	Fc epsilon RI signaling pathway	2.90E-07	2.92E-04
hsa04910	Insulin signaling pathway	1.98E-06	0.001987305
hsa04662	B cell receptor signaling pathway	3.31E-06	0.003332627
hsa04670	Leukocyte transendothelial migration	6.97E-06	0.007011036
hsa04810	Regulation of actin cytoskeleton	1.29E-05	0.012947745
hsa04062	Chemokine signaling pathway	2.86E-05	0.028721255

FDR = false-discovery rate.

DISCUSSION

The IVD is a heterogeneous structure that contributes to load support and flexibility in the spine (Chen et al., 2002); however, it has limited intrinsic capacity for repair (Baer et al., 2001). Therefore, treatments of disc degeneration to date have revolved largely around pain-control measures or spinal surgeries. Alternative approaches such as genetic therapies to slow or reverse the disc degeneration process are urgently needed. In this study, we analyzed the gene expression profile of grade II through IV disc degeneration using bioinformatics. We identified 2 significant DEGs between grade II and III discs and 8 significant DEGs between grade II and IV discs. By constructing an interactive network of the DEGs, we found that MAP and Rho family genes were hub nodes in the network of grade III and IV discs, respectively.

MAP2K6, also known as *MKK6*, encodes a member of the dual-specificity protein kinase family, which functions as an MAP2K. This protein phosphorylates and activates p38 MAPK in response to inflammatory cytokines or environmental stress and plays an important role in many cellular processes such as stress-induced cell cycle arrest, transcription activation, and apoptosis (Kyriakis and Avruch, 2001; Pearson et al., 2001; Kaminska, 2005). Disc degeneration is typically characterized by an imbalance between anabolic and catabolic pro-

cesses and inflammatory mechanisms. Pratsinis and Kletsas (2008) found that growth factors such as insulin-like growth factor-I or basic fibroblast growth factor, which are known to be overexpressed in degenerated disc tissue, can stimulate MAPK and subsequent DNA synthesis in bovine annulus fibrosus and nucleus pulposus cells *in vitro*, suggesting that MAPKs are involved in catabolic and anabolic processes in the IVD.

In addition, MAPK has recently been implicated in the activation of Wnt/β-catenin signals, which may contribute to the pathogenesis of disc degeneration (Hiyama et al., 2011). The MAPK signaling pathways seem to play a crucial role in modulating both matrix synthesis and degradation in the IVD by altering the expression of anabolic and catabolic genes as well as by influencing proteoglycan degradation (Wuertz et al., 2012). In our study, *MAP2K6* was differentially expressed between grade II and III discs, suggesting that this gene plays an important role in the initiation of disc degeneration and therefore may be useful as a specific therapeutic molecular target for treatment.

RHOBTB2 belongs to the Ras homologous (Rho) subfamily, which consists of low-molecular-weight guanosine-5'-triphosphate-binding proteins. Rho proteins are regulatory molecules that link surface receptors to the organization of the actin cytoskeleton and, as such, they mediate changes in cell shape, contractility, motility, and gene expression (Berken and Wittinghofer, 2008; Nowak et al., 2008). Freeman et al. (2008) found that *RHOBTB2* expression is highly upregulated during mitosis, and overexpression of *RHOBTB2* induces a short-term increase in cell cycle progression and proliferation, whereas long-term expression has a negative effect on these processes. Similarly, microarray-based network analysis approaches have found that alteration of *RHOBTB2* levels influences pathways responsible for the cell cycle, apoptosis, cytoskeleton, and membrane-trafficking (Siripurapu et al., 2005). Our findings suggest that *RHOBTB2* is differentially expressed in grade IV compared with grade II discs. However, no data are available about the expression pattern of the RHOB protein in human disc degeneration to date.

We identified 25 dysfunctional pathways in the progression of disc degeneration, and most were associated with preinflammatory cytokine signaling, which is now well recognized during IVD degeneration (Le Maitre et al., 2005, 2007c). Some of the identified dysfunctional pathways are consistent with our knowledge of disc degeneration, and some suggest valuable alternative disc degeneration mechanisms. For example, the neurotrophin signaling pathway is dysfunctional in the progression from grade II to grade III discs. Neurotrophins are signaling molecules involved in the survival, differentiation, migration, and outgrowth of neurons. They are expressed in non-neuronal tissues, including the IVD (Sommer and Kress, 2004), and are beginning to be viewed as important agents therein. Aoki et al. (2004) suggested that nerve growth factor-dependent neurons may be the population responsible for discogenic pain based on findings from studies of disc degeneration in rats. Gruber et al. (2008) found that gene expression levels of brain-derived neurotrophic factor are significantly positively correlated with increasing levels of IVD degeneration.

In conclusion, we analyzed gene expression profiles at different stages of disc degeneration using bioinformatics and found that MAP and Rho family genes - in particular, *MAP2K6* and *RHOBTB2* - may play important roles in the progression of grade III and IV discs, respectively. Therefore, they may be useful as specific therapeutic molecular targets for minimizing toxic side effects in the treatment of disc degeneration. Further experiments are needed to confirm these results.

REFERENCES

Alibés A, Yankilevich P, Cañda A and Díaz-Uriarte R (2007). IDconverter and IDClight: conversion and annotation of gene and protein IDs. *BMC Bioinformatics* 8: 9.

Aoki Y, Ohtori S, Takahashi K, Ino H, et al. (2004). Innervation of the lumbar intervertebral disc by nerve growth factor-dependent neurons related to inflammatory pain. *Spine* 29: 1077-1081.

Baer AE, Wang JY, Kraus VB and Setton LA (2001). Collagen gene expression and mechanical properties of intervertebral disc cell-alginate cultures. *J. Orthop. Res.* 19: 2-10.

Benjamini YHY (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B* (*Methodological*) 57: 289-300.

Berken A and Wittinghofer A (2008). Structure and function of Rho-type molecular switches in plants. *Plant Physiol. Biochem.* 463: 380-393.

Chen J, Baer AE, Paik PY, Yan W, et al. (2002). Matrix protein gene expression in intervertebral disc cells subjected to altered osmolarity. *Biochem. Biophys. Res. Commun.* 293: 932-938.

Doita M, Kanatani T, Harada T and Mizuno K (1996). Immunohistologic study of the ruptured intervertebral disc of the lumbar spine. *Spine* 21: 235-241.

Fraser RD, Osti OL and Vernon-Roberts B (1993). Intervertebral disc degeneration. *Eur. Spine J.* 1: 205-213.

Freeman SN, Ma Y and Cress WD (2008). RhoBTB2 (DBC2) is a mitotic E2F1 target gene with a novel role in apoptosis. *J. Biol. Chem.* 283: 2353-2362.

Freemont AJ, Peacock TE, Goupille P, Hoyland JA, et al. (1997). Nerve ingrowth into diseased intervertebral disc in chronic back pain. *Lancet* 350: 178-181.

Gautier L, Cope L, Bolstad BM and Irizarry RA (2004). affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20: 307-315.

Gruber HE, Ingram JA, Hoelscher G, Zinchenko N, et al. (2008). Brain-derived neurotrophic factor and its receptor in the human and the sand rat intervertebral disc. *Arthritis Res. Ther.* 10: R82.

Gruber HE, Ingram JA, Hoelscher GL, Zinchenko N, et al. (2009). Asporin, a susceptibility gene in osteoarthritis, is expressed at higher levels in the more degenerate human intervertebral disc. *Arthritis Res. Ther.* 11: R47.

Hadjipavlou AG, Tzermiadianos MN, Bogduk N and Zindrick MR (2008). The pathophysiology of disc degeneration: a critical review. *J. Bone Joint Surg. Br.* 90: 1261-1270.

Hiyama A, Sakai D, Tanaka M, Arai F, et al. (2011). The relationship between the Wnt/beta-catenin and TGF-beta/BMP signals in the intervertebral disc cell. *J. Cell Physiol.* 226: 1139-1148.

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

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264.

Kaminska B (2005). MAPK signalling pathways as molecular targets for anti-inflammatory therapy - from molecular mechanisms to therapeutic benefits. *Biochim. Biophys. Acta* 1754: 253-262.

Kyriakis JM and Avruch J (2001). Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol. Rev.* 81: 807-869.

Le Maitre CL, Freemont AJ and Hoyland JA (2004). Localization of degradative enzymes and their inhibitors in the degenerate human intervertebral disc. *J. Pathol.* 204: 47-54.

Le Maitre CL, Freemont AJ and Hoyland JA (2005). The role of interleukin-1 in the pathogenesis of human intervertebral disc degeneration. *Arthritis Res. Ther.* 7: R732-R745.

Le Maitre CL, Freemont AJ and Hoyland JA (2006). Human disc degeneration is associated with increased MMP 7 expression. *Biotech. Histochem.* 81: 125-131.

Le Maitre CL, Freemont AJ and Hoyland JA (2007a). Accelerated cellular senescence in degenerate intervertebral discs: a possible role in the pathogenesis of intervertebral disc degeneration. *Arthritis Res. Ther.* 9: R45.

Le Maitre CL, Hoyland JA and Freemont AJ (2007b). Catabolic cytokine expression in degenerate and herniated human intervertebral discs: IL-1beta and TNFalpha expression profile. *Arthritis Res. Ther.* 9: R77.

Le Maitre CL, Pockert A, Buttle DJ, Freemont AJ, et al. (2007c). Matrix synthesis and degradation in human intervertebral disc degeneration. *Biochem. Soc. Trans.* 35: 652-655.

Luoma K, Riihimaki H, Luukkonen R, Raininko R, et al. (2000). Low back pain in relation to lumbar disc degeneration. *Spine* 25: 487-492.

Nachemson A, Lewin T, Maroudas A and Freeman MA (1970). *In vitro* diffusion of dye through the end-plates and the annulus fibrosus of human lumbar inter-vertebral discs. *Acta Orthop. Scand.* 41: 589-607.

Nagano T, Yonenobu K, Miyamoto S, Tohyama M, et al. (1995). Distribution of the basic fibroblast growth factor and its receptor gene expression in normal and degenerated rat intervertebral discs. *Spine* 20: 1972-1978.

Nowak JM, Grzanka A, Zuryn A and Stepien A (2008). The Rho protein family and its role in the cellular cytoskeleton. *Postepy Hig. Med. Dosw.* 62: 110-117.

Pearson G, Robinson F, Beers GT, Xu BE, et al. (2001). Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr. Rev.* 22: 153-183.

Peng B, Hao J, Hou S, Wu W, et al. (2006). Possible pathogenesis of painful intervertebral disc degeneration. *Spine* 31: 560-566.

Pratsinis H and Kletsas D (2008). Growth factors in intervertebral disc homeostasis. *Connect. Tissue Res.* 49: 273-276.

Shannon P, Markiel A, Ozier O, Baliga NS, et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498-2504.

Siripurapu V, Meth J, Kobayashi N and Hamaguchi M (2005). DBC2 significantly influences cell-cycle, apoptosis, cytoskeleton and membrane-trafficking pathways. *J. Mol. Biol.* 346: 83-89.

Sommer C and Kress M (2004). Recent findings on how proinflammatory cytokines cause pain: peripheral mechanisms in inflammatory and neuropathic hyperalgesia. *Neurosci. Lett.* 361: 184-187.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39: D561-D568.

Team RDC (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Thompson JP, Pearce RH, Schechter MT, Adams ME, et al. (1990). Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. *Spine* 15: 411-415.

Thompson JP, Oegema TR Jr and Bradford DS (1991). Stimulation of mature canine intervertebral disc by growth factors. *Spine* 16: 253-260.

Urban JP and Roberts S (2003). Degeneration of the intervertebral disc. *Arthritis Res. Ther.* 5: 120-130.

van der Laan MJ, Dudoit S and Pollard KS (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat. Appl. Genet. Mol. Biol.* 3: Article15.

Wuertz K, Vo N, Kletsas D and Boos N (2012). Inflammatory and catabolic signalling in intervertebral discs: the roles of NF-kappaB and MAP kinases. *Eur. Cell Mater.* 23: 103-119.