



## Text-mining network analysis of the response to osmotic stimuli in the intervertebral disc

X. Xu\*, L. Liu\* and Q.Y. Lu

Department of Trauma Surgery, East Hospital Affiliated to Tongji University, Shanghai, China

\*These authors contributed equally to this study.

Corresponding author: Q.Y. Lu

E-mail: luqingyoulqy@163.com

Genet. Mol. Res. 12 (2): 1574-1581 (2013)

Received August 10, 2012

Accepted November 21, 2012

Published May 13, 2013

DOI <http://dx.doi.org/10.4238/2013.May.13.11>

**ABSTRACT.** Intervertebral disc cells experience a broad range of physical stimuli under physiologic conditions, including alterations in their osmotic environment. The purpose of this study was to construct a text-mining network of the genes induced during the response to osmotic stimuli in the intervertebral disc. We obtained a gene expression profile of human intervertebral disc cells from the National Center for Biotechnology Information, after culture under hyper- and hypo-osmotic conditions compared to iso-osmotic conditions, and we identified 65 differentially expressed genes of intervertebral disc cells. We constructed a text-mining network using Biblio-MetReS between the differentially expressed genes and other genes that were included in the same document as the differentially expressed genes. Then, we performed pathway-enrichment analysis to identify the most relevant pathways for the response to osmotic stimuli in intervertebral disc cells. Our data provide a comprehensive bioinformatics analysis of genes and pathways that may be involved in the response to osmotic stimuli in the intervertebral disc.

**Key words:** Text-mining network; Intervertebral disc; Osmotic stimuli

## INTRODUCTION

The intervertebral disc (IVD) is an avascular and alymphatic tissue that consists of three morphologically distinct zones: nucleus pulposus, annulus fibrosus, and transition zone (Li et al., 2008). The IVD is very important in maintaining the flexibility of the spine and dissipating applied loads. The extracellular matrix of the IVD is comprised largely of water (60 to 99% by weight) and negatively charged proteoglycans. Therefore, IVD cells exposed to physiological mechanical loads experience complex physical stimuli including strain, fluid flow, osmotic and hydrostatic pressure, and compressive, tensile, and shear stresses (Weidenbaum et al., 1992; Urban et al., 1993; Iatridis et al., 1996). IVD cells can secrete a complex extracellular matrix that contains the highly sulfated proteoglycan aggrecan through its interactions with cations and provides the tissue with its unique osmotic properties (Cheng et al., 2009).

Reconstructing molecular networks that are responsible for regulating biological processes is a fundamental task in molecular biology (McIntosh and Curran, 2009). Various biological problems have already been successfully addressed using text mining of the biomedical literature in recent years, including the discovery and characterization of protein sorting (Craven and Kumlien, 1999; Stapley et al., 2002), molecular binding (Liu et al., 2005), and molecular interactions, e.g., protein-protein (Ng and Wong, 1999; Wong, 2001; Park et al., 2001; Yakushiji et al., 2001; Friedman et al., 2001), gene-protein (Sekimizu et al., 1998), and gene-drug (Rindfleisch et al., 2000).

Taking the large number of available documents into consideration, it is difficult to mine all of the available information in order to reconstruct networks. One method relies on the automated analysis of published scientific literature that is available in public databases over the Internet to identify genes and proteins that are found in the same document(s) (Hoffmann and Valencia, 2004, 2005; Hahn and Valencia, 2006; von Mering et al., 2007; Krallinger et al., 2008). Biblio-MetReS is one tool that was proposed in order to reconstruct the network of genes associated with a given gene of interest using automated mining of the published literature. Biblio-MetReS creates networks that are comparable to those of other user friendly tools. Furthermore, it permits analysis of the full text of scientific documents, which provides more complete reconstructions than from those using only the abstract of a document (Usie et al., 2011).

In the present study, we identified the differentially expressed genes (DEGs) of IVD cells following culture under hyper- and hypo-osmotic conditions compared to iso-osmotic conditions, and constructed a text-mining network using Biblio-MetReS between the DEGs and other genes that were found in the same document as the DEGs. We then performed pathway-enrichment analysis to identify the most relevant pathways in the response to osmotic stimuli in the IVD. We aimed to identify the genes and pathways that are regulated by osmotic conditions in IVD cells.

## MATERIAL AND METHODS

### Microarray data

The transcription profile GSE1648 (Boyd et al., 2005) was obtained from the public

functional genomics data repository Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>), which is based on the Affymetrix Human Genome U133A Array. A total 11 samples, including 4 hyper-osmotic stimuli, 4 iso-osmotic stimuli, and 3 hypo-osmotic stimuli, were applied to identify DEGs. These three groups were used to investigate the response of the IVD to osmotic stimuli.

### Construction of the text-mining network

In order to construct a text-mining network, we used Biblio-MetReS to perform the analysis. Firstly, we registered with Biblio-MetReS in order to access the central database. Then, *Homo sapiens* was chosen as the target species. After inputting the DEGs and selecting the literature database (MEDLINE, PubMed, and BioMed Central), the search was started. The search tool identifies those documents that contain the target gene names and their synonyms. It then extracts the full text from each document and searches for the co-occurrence of any pair of genes. Finally, we extracted the co-occurrence of any pair of DEGs from Biblio-MetReS and used Cytoscape (Smoot et al., 2011) to display the text-mining network.

### Pathway-enrichment analysis

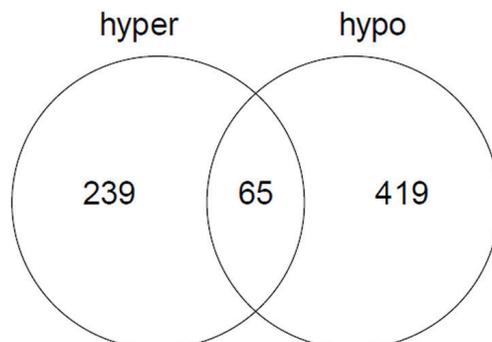
The Kyoto Encyclopedia of Genes and Genomes is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals (Kanehisa, 2002). The PATHWAY database records the networks of intracellular molecular interactions and their variants in specific organisms (<http://www.genome.jp/kegg/>).

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al., 2009), which is a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. We used DAVID to identify over-represented pathways using the genes isolated from the text-mining network. Pathways with a false-discovery rate (Benjamini and Yekutieli, 2005) less than 0.1 and a count larger than 2 were selected.

## RESULTS

### Selection of DEGs

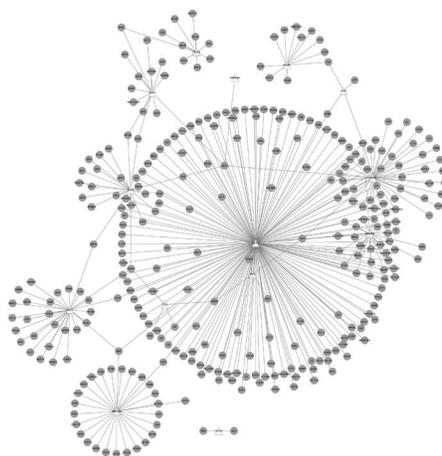
In order to identify DEGs, we obtained the publicly available microarray dataset GSE1648 from GEO. For the GSE1648 dataset, the limma method (Smyth, 2004) was used to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the Robust Multichip Average method with the default settings implemented in Bioconductor, and then constructed a linear model. After microarray analysis, DEGs with a fold-change larger than 2 and a P less than 0.05 were selected. In all, 304 genes were identified in IVD cells as significantly differentially expressed between hyper- and iso-osmotic stimuli specimens, while a total of 484 genes were significantly differentially expressed between hypo- and iso-osmotic stimuli specimens. A total of 65 overlapping genes were selected as DEGs from GSE1648 (Figure 1).



**Figure 1.** VENN graph display the information of our data. A total of 65 overlapping genes were selected as differentially expressed genes from GSE1648.

### Construction of the text-mining network

To identify the relationships between DEGs and other genes, we used Biblio-MetReS to construct a text-mining network. After removing the genes that were not related to the DEGs, we were left with 14 DEGs and 317 other genes. By integrating the relationships described above, a text-mining network of the genes induced in response to osmotic stimuli in the IVD was built (Figure 2). In the text-mining network, cystic fibrosis transmembrane conductance regulator (*CFTR*), neuregulin 1 (*NRG1*), insulin-like growth factor (*IGF1*), and *CD28* were “hub” genes, suggesting that these genes play a crucial role in the response to osmotic stimuli in the IVD. *CFTR*, as a hub gene, was linked to a large number of target genes, e.g., *SHANK*, *CD74*, and interferon alpha (*IFNA*). In addition, our network revealed that SMAD family member 4 alpha (*SMAD4A*) and interferon regulatory factor 5 (*IRF5*) were both linked to *CD28* by interacting indirectly with *CD74* and *IFNA1*.



**Figure 2.** Text-mining network of gene responses to osmotic stimuli in intervertebral disc. The nodes represent differentially expressed genes and the other nodes represent other genes. The genes (nodes), which are connected by lines, suggest that they were connected in the same literature.

## Pathway-enrichment analysis

To identify the relevant pathways that changed in response to osmotic stimuli in the IVD, we used DAVID to perform pathway-enrichment analysis. In order to focus on the most significantly relevant pathways, a false-discovery rate less than 0.1 and a count larger than 2 were chosen as the threshold. Pathway-enrichment analysis yielded many significant pathways, which included “Toll-like receptor signaling pathway”, “Autoimmune thyroid disease”, and “Focal adhesion” (Table 1). Phosphoinositide-3-kinase (PI3K), catalytic, gamma polypeptide (*PIK3CG*) was enriched in the “Toll-like receptor signaling pathway”, “Aldosterone-regulated sodium reabsorption”, “Focal adhesion”, and “T cell receptor signaling pathway”. Interleukin 6 (*IL-6*) was enriched in the “Toll-like receptor signaling pathway”, “Pathways in cancer”, and “Intestinal immune network for IgA production”.

**Table 1.** Pathway-enrichment analysis.

Category	Term	Gene count	FDR
hsa04620	Toll-like receptor signaling pathway	19	8.59E-05
hsa04960	Aldosterone-regulated sodium reabsorption	11	0.004026
hsa05320	Autoimmune thyroid disease	12	0.00475
hsa04510	Focal adhesion	24	0.00525
hsa04660	T cell receptor signaling pathway	17	0.006091
hsa05200	Pathways in cancer	32	0.006435
hsa04672	Intestinal immune network for IgA production	11	0.022239
hsa05330	Allograft rejection	9	0.08184

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). IVD cells experience a broad range of physical stimuli under physiologic conditions, including alterations in their osmotic environment (Boyd et al., 2005). The recent development of Biblio-MetReS, a tool to reconstruct the network of genes associated with a given gene by automated mining of the published literature, has made it possible to mine all of the available information in order to reconstruct networks. To date, text-mining of the genes induced in response to osmosis stimuli in the IVD has not been performed. Therefore, in this study, we obtained the gene expression profile of human IVD cells from the National Center for Biotechnology Information, and identified the DEGs of IVD cells following culture under hyper- and hypo-osmotic conditions compared to iso-osmotic conditions. We constructed a text-mining network using Biblio-MetReS between the DEGs and other genes that were found in the same document as the DEGs. We then performed pathway-enrichment analysis to identify the most relevant pathways in the response to osmotic stimuli in the IVD. We aimed to identify genes and pathways that are regulated by osmotic conditions in IVD cells.

Using text-mining network construction of the IVD, we identified many genes associated with the response to osmotic stimuli. *CFTR*, *NRG1*, *IGF1*, and *CD28* acted as hub genes that were linked to a large number of target genes.

*CFTR* belongs to the ATP-binding cassette transporter superfamily and is involved in transport, both physiologically and pathologically, in various systems (Goodman and Percy,

2005). It functions as a channel across the membrane of cells that produce mucus, sweat, saliva, tears, and digestive enzymes. The channel transports negatively charged chloride ions in and out of cells, and the transport of chloride helps to control the movement of water (osmosis) in tissues and maintains the fluidity of mucus and other secretions (Taddei et al., 2004). Our result confirmed that CFTR plays an important role in the response to osmotic stimuli in the IVD.

PIK3CG encodes a protein that belongs to the PI3K/PI4K family of proteins. It is an important modulator of extracellular signals, including those elicited by E-cadherin-mediated cell-cell adhesion, which plays an important role in the maintenance of the structural and functional integrity of epithelial tissues (Laprise et al., 2002). In a recent study, Cheng et al. (2009) reported that the PI3K/AKT signaling pathway plays an important role in the regulation of aggrecan gene expression in cells of the nucleus pulposus in IVD. Importantly, aggrecan gene expression appears to change in chondrocytes in response to hypo- and hyper-osmotic stimuli (Chen et al., 2002), and IVD cells can secrete a complex extracellular matrix that contains the highly sulfated proteoglycan aggrecan that provides the tissue with its unique osmotic properties (Cheng et al., 2009). IL-6 has multiple functions in a wide variety of biological processes, including the immune system, neuronal regulation, hematopoiesis, and inflammation (Gadient and Otten, 1997; Barnes et al., 2011). Relevant to this study are the findings that a protruded IVD was immunoreactive for IL-6 at a 4-fold higher intensity than normal tissue (Specchia et al., 2002). Therefore, this cytokine may have an important, but unknown, function in regulating cell function in IVD pathology.

Many significant pathways were identified in our analysis, e.g., “Toll-like receptor signaling pathway”, “Autoimmune thyroid disease”, and “Focal adhesion”. Transforming growth factor activated kinase 1 (TAK1) plays an essential role in Toll-like receptor signaling pathways (Shim et al., 2005). In addition, several studies (Shirakabe et al., 1997; Cheung et al., 2003; Singhirunnusorn et al., 2005) have shown that TAK1 can be activated by chemical and physical stresses, including osmotic stress, suggesting that the “Toll-like receptor signaling pathway” is likely involved in the response to osmotic stimuli in the IVD. Focal adhesions play essential roles in many important biological processes including cell motility, cell proliferation, cell differentiation, regulation of gene expression, and cell survival. Some of the constituents of focal adhesions participate in the structural link between membrane receptors and the actin cytoskeleton, while others are signaling molecules, including different protein kinases and phosphatases, their substrates, and various adapter proteins. Our findings suggest that changes in the osmotic environment of IVD cells may modify the expression of proteins involved in focal adhesion.

In conclusion, our data provide a comprehensive bioinformatics analysis of genes and pathways that may be involved in the response to osmotic stimuli in the IVD. We identified 65 DEGs from GSE1648, and constructed a text-mining network using Biblio-MetReS. Furthermore, eight pathways were involved in the response to osmotic stimuli in our pathway-enrichment analysis. However, additional insights into the gene expression changes in IVD cells regulated by osmotic stimuli are still needed to confirm our findings.

## REFERENCES

- Barnes TC, Anderson ME and Moots RJ (2011). The many faces of interleukin-6: the role of IL-6 in inflammation, vasculopathy, and fibrosis in systemic sclerosis. *Int. J. Rheumatol.* 2011: 721608.
- Benjamini Y and Yekutieli D (2005). Quantitative trait Loci analysis using the false discovery rate. *Genetics* 171: 783-790.

- Boyd LM, Richardson WJ, Chen J, Kraus VB, et al. (2005). Osmolarity regulates gene expression in intervertebral disc cells determined by gene array and real-time quantitative RT-PCR. *Ann. Biomed. Eng.* 33: 1071-1077.
- 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.
- Cheng CC, Uchiyama Y, Hiyama A, Gajghate S, et al. (2009). PI3K/AKT regulates aggrecan gene expression by modulating Sox9 expression and activity in nucleus pulposus cells of the intervertebral disc. *J. Cell Physiol.* 221: 668-676.
- Cheung PC, Campbell DG, Nebreda AR and Cohen P (2003). Feedback control of the protein kinase TAK1 by SAPK2a/p38alpha. *EMBO J.* 22: 5793-5805.
- Craven M and Kumlien J (1999). Constructing biological knowledge bases by extracting information from text sources. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 77-86.
- Friedman C, Kra P, Yu H, Krauthammer M, et al. (2001). GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics* 17 (Suppl 1): S74-S82.
- Gadient RA and Otten UH (1997). Interleukin-6 (IL-6) - a molecule with both beneficial and destructive potentials. *Prog. Neurobiol.* 52: 379-390.
- Goodman BE and Percy WH (2005). CFTR in cystic fibrosis and cholera: from membrane transport to clinical practice. *Adv. Physiol. Educ.* 29: 75-82.
- Hahn U and Valencia A (2006). Semantic Mining in Biomedicine (Introduction to the papers selected from the SMBM 2005 Symposium, Hinxton, U.K., April 2005). *Bioinformatics* 22: 643-644.
- Hoffmann R and Valencia A (2004). A gene network for navigating the literature. *Nat. Genet.* 36: 664.
- Hoffmann R and Valencia A (2005). Implementing the iHOP concept for navigation of biomedical literature. *Bioinformatics* 21 (Suppl 2): ii252-ii258.
- Huang dW, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.
- Iatridis JC, Weidenbaum M, Setton LA and Mow VC (1996). Is the nucleus pulposus a solid or a fluid? Mechanical behaviors of the nucleus pulposus of the human intervertebral disc. *Spine* 21: 1174-1184.
- Kanehisa M (2002). The KEGG database. *Novartis Found. Symp.* 247: 91-101.
- Krallinger M, Valencia A and Hirschman L (2008). Linking genes to literature: text mining, information extraction, and retrieval applications for biology. *Genome Biol.* 9 (Suppl 2): S8.
- Laprise P, Chailier P, Houde M, Beaulieu JF, et al. (2002). Phosphatidylinositol 3-kinase controls human intestinal epithelial cell differentiation by promoting adherens junction assembly and p38 MAPK activation. *J. Biol. Chem.* 277: 8226-8234.
- Li S, Duance VC and Blain EJ (2008). Zonal variations in cytoskeletal element organization, mRNA and protein expression in the intervertebral disc. *J. Anat.* 213: 725-732.
- Liu Y, Navathe SB, Civera J, Dasigi V, et al. (2005). Text mining biomedical literature for discovering gene-to-gene relationships: a comparative study of algorithms. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2: 62-76.
- McIntosh T and Curran JR (2009). Challenges for automatically extracting molecular interactions from full-text articles. *BMC Bioinformatics* 10: 311.
- Ng SK and Wong M (1999). Toward routine automatic pathway discovery from on-line scientific text abstracts. *Genome Inform. Ser. Workshop Genome Inform.* 10: 104-112.
- Park JC, Kim HS and Kim JJ (2001). Bidirectional incremental parsing for automatic pathway identification with combinatory categorial grammar. *Pac. Symp. Biocomput.* 396-407.
- Rindfleisch TC, Tanabe L, Weinstein JN and Hunter L (2000). EDGAR: extraction of drugs, genes and relations from the biomedical literature. *Pac. Symp. Biocomput.* 517-528.
- Sekimizu T, Park HS and Tsujii J (1998). Identifying the interaction between genes and gene products based on frequently seen verbs in medline abstracts. *Genome Inform. Ser. Workshop Genome Inform.* 9: 62-71.
- Shim JH, Xiao C, Paschal AE, Bailey ST, et al. (2005). TAK1, but not TAB1 or TAB2, plays an essential role in multiple signaling pathways *in vivo*. *Genes Dev.* 19: 2668-2681.
- Shirakabe K, Yamaguchi K, Shibuya H, Irie K, et al. (1997). TAK1 mediates the ceramide signaling to stress-activated protein kinase/c-Jun N-terminal kinase. *J. Biol. Chem.* 272: 8141-8144.
- Singhirunnusorn P, Suzuki S, Kawasaki N, Saiki I, et al. (2005). Critical roles of threonine 187 phosphorylation in cellular stress-induced rapid and transient activation of transforming growth factor-beta-activated kinase 1 (TAK1) in a signaling complex containing TAK1-binding protein TAB1 and TAB2. *J. Biol. Chem.* 280: 7359-7368.
- Smoot ME, Ono K, Ruscheinski J, Wang PL, et al. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431-432.
- Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray

- experiments. *Stat. Appl. Genet. Mol. Biol.* 3: Article3.
- Specchia N, Pagnotta A, Toesca A and Greco F (2002). Cytokines and growth factors in the protruded intervertebral disc of the lumbar spine. *Eur. Spine J.* 11: 145-151.
- Stapley BJ, Kelley LA and Sternberg MJ (2002). Predicting the sub-cellular location of proteins from text using support vector machines. *Pac. Symp. Biocomput.* 374-385.
- Taddei A, Folli C, Zegarra-Moran O, Fanen P, et al. (2004). Altered channel gating mechanism for CFTR inhibition by a high-affinity thiazolidinone blocker. *FEBS Lett.* 558: 52-56.
- Urban JP, Hall AC and Gehl KA (1993). Regulation of matrix synthesis rates by the ionic and osmotic environment of articular chondrocytes. *J. Cell Physiol.* 154: 262-270.
- Usie A, Karathia H, Teixeira I, Valls J, et al. (2011). Biblio-MetReS: a bibliometric network reconstruction application and server. *BMC Bioinformatics* 12: 387.
- von Mering C, Jensen LJ, Kuhn M, Chaffron S, et al. (2007). STRING 7 - recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* 35: D358-D362.
- Weidenbaum M, Foster RJ, Best BA, Saed-Nejad F, et al. (1992). Correlating magnetic resonance imaging with the biochemical content of the normal human intervertebral disc. *J. Orthop. Res.* 10: 552-561.
- Wong L (2001). PIES, a protein interaction extraction system. *Pac. Symp. Biocomput.* 520-531.
- Yakushiji A, Tateisi Y, Miyao Y and Tsujii J (2001). Event extraction from biomedical papers using a full parser. *Pac. Symp. Biocomput.* 408-419.