European Review for Medical and Pharmacological Sciences 2013; 17: 618-623

The PPI network and cluster ONE analysis to explain the mechanism of bladder cancer

F.-C. WAN, Y.-P. CUI, J.-T. WU, J.-M. WANG, Q.-Z LIU, Z.-L. GAO

Department of Urology, Yantai Yuhuangding Hospital, Yantai, Shandong Province, China.
Fengchun Wan and Yupeng Cui should be regarded as co-first Authors

Abstract. – BACKGROUND: Bladder cancer is a common cancer worldwide whose incidence continues to increase. It is estimated that there are 261,000 cases of bladder cancer resulting in 115,000 deaths worldwide.

AIM: Although some studies can be initiated using small local tissue collections, high quality collection of fresh tissues from new clinical trials will be crucial for proper evaluation of associations with clinical outcome. For superficial bladder cancer, identification of tumors that will progress has long been perceived as a potential application of genetic studies.

MATERIALS AND METHODS: In our study, we constructed the Protein-Protein Interactions (PPI) network using the Cytoscape and detected some network modeling clusters. In addition, we enriched GO categories among these genes in the first cluster and detected a pathway i.e. Spliceosome (hsa03040). Most Gene Ontology (GO) categories and Spliceosome were closely to RNA splicing and cellular macromolecular complex (CMC) assembly, which indicates that the mutation of RNA splicing and CMC assembly maybe important factors causing bladder cancer.

RESULTS: In our study, these clusters of GO:0034622, GO:0006397 and GO:0034621 in bladder cancer belong to cellular macromolecular complex assembly, which may play an important role in the occurrence of cancer cells.

CONCLUSIONS: It is a great significance for the detection and treatment of bladder cancer to understand the mechanism of RNA splicing and CMC assembly.

Key Words:
Bladder cancer, PPI network, Network modeling, GO categories, Bio-pathway.

Introduction

Bladder cancer is the fifth most common cancer and caused more than 115,000 deaths in the world in[1] 2008ett. Ninety-nine percent of bladder cancers are transitional cell carcinoma. The other 10% are squamous cell carcinoma, adenocarcinoma, sarcoma, small cell carcinoma, and secondary deposits from cancers elsewhere in the body. These tumors originate in the bladder mucosa, progressively invade the lamina propria, and move sequentially into the muscularispropria, perivesical fat, and contiguous pelvic structures, with increasing incidence of lymph node involvement with progression[2,3].

Bladder cancer belongs to transitional cell carcinoma (TCC). TCCs are often multifocal, with 30-40% of patients having more than one tumor at diagnosis. The pattern of growth of blander cancer can be papillary, sessile (flat) or carcinoma-in-situ (CIS). Although some studies can be initiated using small local tissue collections, high quality collection of fresh tissues from new clinical trials will be crucial for proper evaluation of associations with clinical outcome[4,5]. For superficial bladder cancer, identification of tumors that will progress has long been perceived as a potential application of genetic studies[6]. Because bladder cancer is often multifocal and some tumors arise within a urothelium that shows widespread urothelialatypia, objective methods are needed to assess the status of the urothelium remaining after resection of all overt tumors, possibly by the assessment of irrigation specimens[7].

In our study, we identified the DEGs using the limma package and constructed the PPT network using the Cytoscape. In addition, we found network clusters in PPI network modeling and further obtained higher over expression clusters by MCODE analysis. Finally, ten GO terms were enriched with GO function annotation and one KEGG pathway, i.e. Spliceosome (hsa03040) were detected.

Data and Methods

Data Source

The transcription profile of GSE3167[8] was obtained from *National Center for Biotechnology Information Gene Expression Omnibus* (NCBI GEO) database (http://www.ncbi.nlm.nih.gov/geo/) which is based on the Affymetrix Human Genome U133A Array. A Total of 60 chips, purchased from

Corresponding Author: Zhenli Gao, MD; e-mail: gaozhenlizhenli@hotmail.com

Department of Clinical Biochemistry in Aarhus University Hospital, Skejby in Denmark, were used for our analysis.

In this study, we used microarray expression profile to examine the gene expression patterns in superficial transitional cell carcinoma (sTCC) with surrounding Carcinoma in situ-(CIS) (13 patients), without surrounding CIS lesions (15 patients) and in muscle invasive carcinomas (mTCC; 13 patients). Hierarchical cluster analysis separated the sTCC samples according to the presence or absence of CIS in the surrounding urothelium. We identified a few gene clusters that contained genes with similar expression levels in TCC with surrounding CIS and invasive TCC. However, no close relationship between TCC with adjacent CIS and invasive TCC was observed using hierarchical cluster analysis. Expression profiling of a series of biopsies from normal urothelium and urothelium with CIS lesions from the same urinary bladder revealed that the gene expression found in sTCC with surrounding CIS was found also in CIS biopsies as well as in normal samples adjacent to the CIS lesions. Furthermore, we also identified similar gene expression changes in mTCC samples. We used a supervised learning approach to build a 16-gene molecular CIS classifier. The classifier was able to classify sTCC samples according to the presence or absence of surrounding CIS with a high accuracy. This study demonstrates that a CIS gene expression signature is present not only in CIS biopsies but also in sTCC, mTCC, and, remarkably, in histologically normal urothelium from bladders with CIS. Identification of this expression signature could provide guidance for the selection of therapy and follow-up regimen in patients with early stage bladder cancer.

The Human Protein Reference Database (HPRD)[9] is a protein database accessible through the internet. The Biological General Repository for Interaction Datasets (BioGRID)[10] is a curated biological database of protein-protein and genetic interactions. Total 326119 unique Protein-Protein *Interaction* (PPI) pairs were collected in which 39240 pairs are from HPRD and 379426 pairs are from BioGRID.

Differentially Expressed Genes (DEGs) Analysis

For the GSE3167 dataset, the limma package[11] was used to identify differentially expressed genes (DEGs). The original expression datasets from all conditions were extracted into expression estimates, and then constructed to the linear model. The DEGs only with the fold change value larger than 2 and p-value less than 0.05 were selected.

Protein-Protein Interaction (PPI) Network Construction

For demonstrating the potential PPI relationship, the Pearson Correlation Coefficient (PCC)[12] was calculated for all pair-wise comparisons of gene-expression values between ordinary genes and the DEGs. The PPI relationships whose absolute PCC are larger than 0.6 were considered as significant.

To further understand the potential PPI relationship, we matched the interactions between two DEGs using the protein interaction (PPI data that have been collected from HPRD and BIOGRID database). Based on the above datasets, PPI network was constructed using the Cytoscape[13].

Network Modeling in Cytoscape

The network is un-weighted, i.e. noscore is assigned to the edges. When a protein was very well studied, lots of experiments would describe its partners and one interaction could be identified several times. As the number of occurrences of an interaction was considered as a criterion of reliability, it could be advantageous to attribute a higher weight to the edges that were the more frequent. However, the weighting would introduce an important bias, for it would favor the most studied proteins.

Cytoscape MCODE Analysis

Cluster with overlapping Neighbourhood Expansion (Cluster ONE)[14] is used to discover densely connected and possibly overlapping regions within the Cytoscape network you have constructed. In PPI networks, these dense regions usually respond to protein complexes or fractions of them. Cluster ONE works by "growing" dense regions out of small seeds guided by a quality function. The quality of a group is evaluated by the number of internal divided edges involving nodes of the group. In a PPI network, subgraphs of highly interconnected proteins can be considered as protein complexes or functional modules. Subgraphs smaller than 3 or having a density less than 0.25 (number of edges within the cluster divided by the number of theoretically possible edges) and p-value < 0.05, were discarded.

When a large cluster cannot be further split into over lapping subgraphs of highly interconnected nodes, it can be split into independent subgraphs of highly interconnected nodes. The partition into independent subgraphs was performed with the plug in Molecular Complex Detection[15] (MCODE) of Cytoscape. This algorithm detects densely connected regions. First it assigned a weight to each node, corresponding to its local neighbourhood density. Then, it recursively moved outward, including in the cluster the nodes whose weight is above a given threshold. The default parameters of MCODE plug in was degree cutoff ≥ 2, node score cutoff ≥ 0.2, k-core ≥ 2.

Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway Analysis

Database for Annotation, Visualization and Integrated Discovery (*DAVID*)[16], a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. In David tool, a cumulative hypergeometric distribution is used for calculating the probability of getting at least n successes in the hypergeometric experiment. A cumulative hypergeometric probability refers to a sum of probabilities associated with a hypergeometric experiment. To compute a cumulative hypergeometric probability, we may need to add one or more individual probabilities.

Formula:

$$P = 21 \subsetneq \frac{\binom{fnf}{imi}\binom{2}{2}}{\binom{n}{m}}, \quad i=0,\ k21$$

In formula, n represents the number of genes in PPI network. F represents the number of proteins, which have GO function annotation. M represents the number of proteins, which involve biology pathways. K represents the frequency of GO-ID (Gene Ontology-Identification) emergency. We used the probability that was set as $p < 0.05$ and count > 2 to identify over-represented GO categories in biological process and KEGG pathway analysis based on the cumulative hypergeometric distribution.

Results

PPI Network Modeling in Cytoscape

Publicly available microarray data set GSE3167are obtained from GEO. Total 1404 DEGs with the fold change value > 2 and *p*-value < 0.05 were selected using the limma package. All of these genes are positive expression genes.

To obtain PPI network, minimum size 4,413expression relationships including 198normal genes and their 169 DEGs were selected. By integrating expression relationships above, a PPI network in Pancreatic Cancer was built between its DEGs and normal genes (Figure 1).

Network Clustering

To find Network clustering, we set basic parameters: minimum size is 9 and minimum density is 0.3. Network clustering was found in PPI Network modeling (Figure 2).

Cytoscape MCODE Analysis

To obtain higher over expression clusters, we analysis network by MCODE, which are taken from Cytoscape.MCODE does not provide any statistical score on the resulting clusters but can be used as a discovery tool in network analysis.

Function Analysis of the First Cluster

Using the GO terms to descript the function of the first cluster (Figure 3), several Gene Ontology (GO) categories were enriched among these genes in the first cluster, including spliceosomal snRNP biogenesis (GO:0000387), ribonucleoprotein complex assembly (GO: 0022618), RNA splicing, via transesterification reactions (GO:0000375) and so on (Table I). With GO function annotation using DAVID tool, top10 enriched GO terms were list in Table I.

Using the KEGG pathways to descript the function of the first cluster (Figure 3), one KEGG pathways were detected in the first cluster, i.e. Spliceosome (hsa03040).

Discussion

Bladder cancer is the second most common genitourinary malignancy, with transitional-cell carcinoma (TCC) comprising nearly 90% of all primary bladder tumors[17]. Although the majority of patients present with superficial bladder tumors, 20% to 40% either present with or develop invasive disease[18]. We constructed a PPI network in Pancreatic Cancer built between its DEGs and

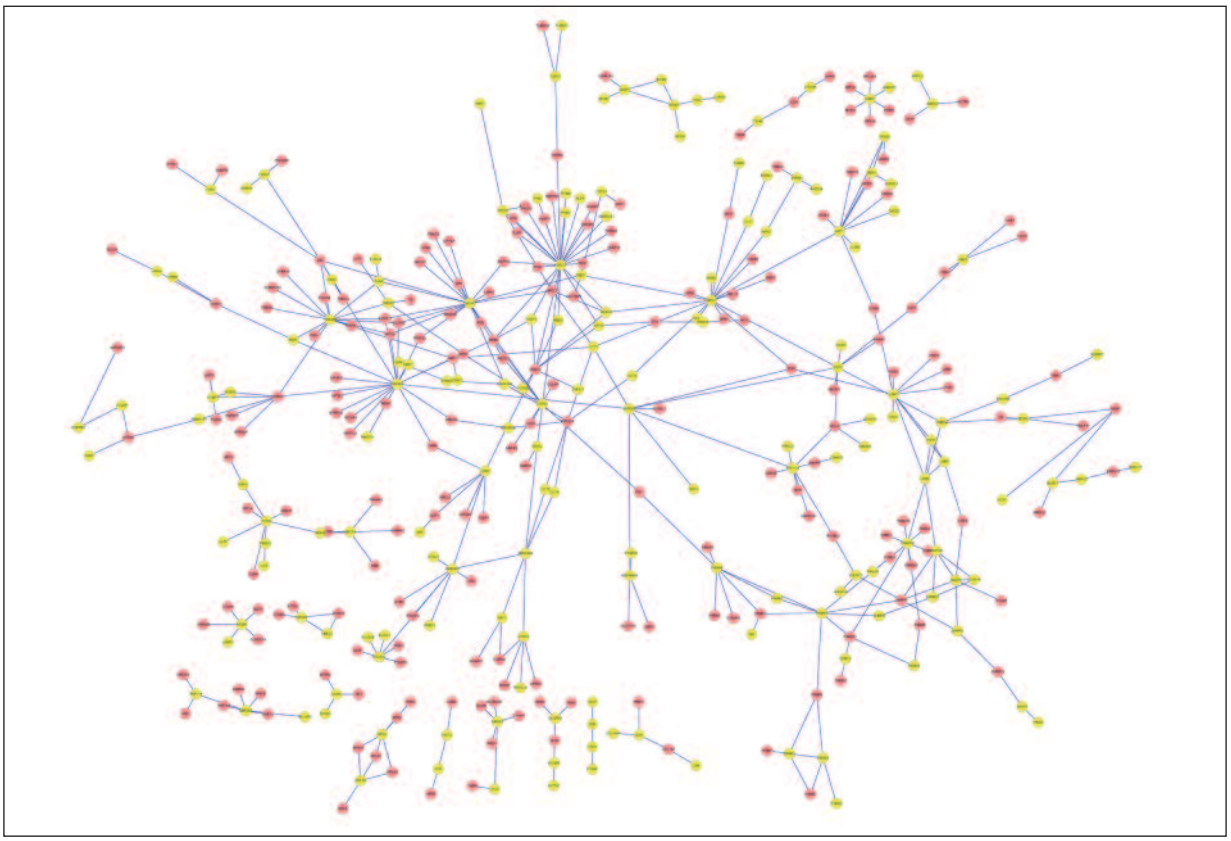

Figure 1. PPI Network modeling in Cytoscape, the yellow nodes represent DEGs and the pink nodes represent normal genes.

normal genes, which were from HPRD and BioGRID database. We analysis PPI network by Cluster ONE and MCODE, which can be used as a discovery tool in network analysis, for it would favor the most studied proteins. Ten enriched GO terms were got with GO functional analysis and Spliceosome (hsa03040) was obtained using the KEGG pathways.

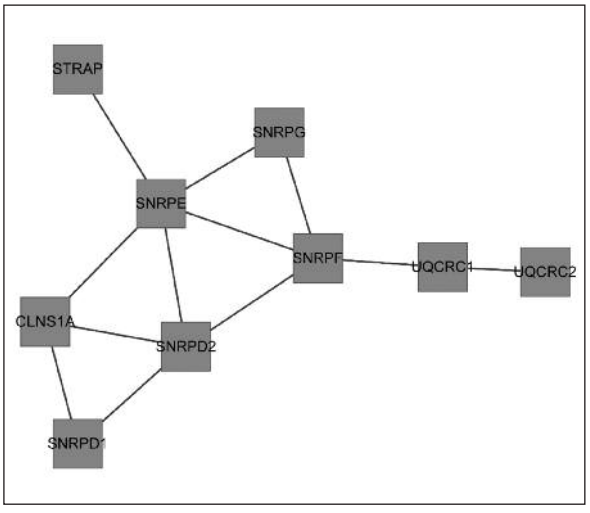

Figure 2. Network clustering. They came from the partition of the whole PPI network modeling with Cluster ONE.

In the past few years, protein-protein interaction (PPI) networks of several organisms have been derived and made publicly available[19]. In the recent literature, two additional measures have been used to compare PPI networks with random network models[20]. In our study, PPI network was constructed using the Cytoscape, which is an open source software project for integrating biomolecular interaction networks with high-throughput expression data and other molecular states into a unified conceptual framework[13]. Protein-protein interaction (PPI) network represent the interactions among proteins in an organism, where each protein is represented with a node, and each interaction is represented with an edge between two nodes. The construction of PPI Network provides a great convenience for our work. Based on PPI network we got network clustering. Furthermore, MCODE was used for over expression clusters, which was detected in Figure 3.

Table I. Top10 enriched GO terms.

Term	Description	Count	p-value	FDR
GO:0000387	Spliceosomal snRNP biogenesis	6	2.60E-14	2.09E-11
GO:0022618	Ribonucleoprotein complex assembly	6	2.98E-12	2.39E-09
GO:0000375	RNA splicing, via transesterification reactions	6	1.73E-10	1.39E-07
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	6	1.73E-10	1.39E-07
GO:0000398	Nuclear mRNA splicing, via spliceosome	6	1.73E-10	1.39E-07
GO:0022613	Ribonucleoprotein complex biogenesis	6	3.95E-10	3.17E-07
GO:0008380	RNA splicing	6	3.94E-09	3.16E-06
GO:0034622	Cellular macromolecular complex assembly	6	6.96E-09	5.58E-06
GO:0006397	mRNA processing	6	7.30E-09	5.85E-06
GO:0034621	Cellular macromolecular complex subunit organization	6	1.25E-08	9.99E-06

The Sm class of small nuclear ribonucleoproteins (snRNPS) is major constituents of the spliceosome, the catalytic center of the pre-mRNA splicing reaction[21]. To date, the only known function for the Sm proteins is in the biogenesis of U snRNPs, and the biogenesis of snRNPs U1, U2, U4, and U5 is a complex cycle that requires the bidirectional transport of these snRNAs across the nuclear envelope[22,23]. Ribonucleoprotein (RNP) is a nucleoprotein that contains RNA, i.e. it is an association that combines ribonucleic acid and protein together. RNP in snRNPs has an RNA-binding motif in its RNA-binding protein. Aromatic amino acid residues in this motif result in stacking interactions with RNA[24]. Significant progress has been made in identifying the components of the splicing machinery and determining the general pathway of the reaction. Less has been learned about the determinants of splice site recognition and selection.

Together with GO:0000375, GO: 0000377, GO: 0000398, GO: 0022613 and GO: 0008380, they all belonged to RNA splicing clusters. RNA splicing is a modification of the nascent pre-mRNA taking place after or concurrently with its transcription, in which introns are removed and exons are joined[21]. This is needed for the typical eukaryotic messenger RNA before it can be used to produce a correct protein through translation[25]. From all the data, we can predict that the development of mechanisms of bladder cancer was closely related to RNA splicing. It is a great significance for the detection and treatment of bladder cancer to understand the mechanism of RNA splicing.

The living cells contain specific macromolecules with a high molecular weight and a poor solubility. Generally in a cell, macromolecules are represented by polysaccharides, proteins, nucleic acids and enzymes. These compounds are formed by polymerisation of micromolecules such as sugars, amino acids and nucleotides. It is a key work to study synthesis mechanism of these macromolecules because they control and regulate proliferation and differentiation of cell in different tissue. In our study, these clusters of GO:0034622, GO:0006397 and GO:0034621 in bladder cancer belong to cellular macromolecular complex assembly, which may play an important role in the occurrence of cancer cells.

In KEGG pathways analysis, we detected only one pathway, i.e. Spliceosome (hsa03040), which is a complex of snRNA and protein subunits that removes introns from a transcribed pre-mRNA (hnRNA) segment. Spliceosome, a multimegadalton ribonucleoprotein (RNP) complex comprised of five snRNPs and numerous proteins, catalyzes pre-mRNA splicing[26]. This further illustrates that the mutation of RNA splicing maybe an important factor causing bladder cancer. Intricate RNA-RNA

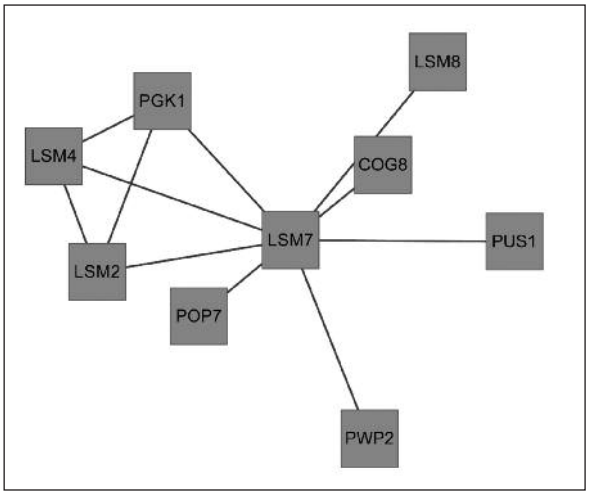

Figure 3. Cytoscape MCODE analysis. A and B came from the partition of network clustering A and B (Figure 2) respectively with MCODE.

and RNP networks, which serve to align the reactive groups of the pre-mRNA for catalysis, are formed and repeatedly rearranged during spliceosome assembly and catalysis[26].

Acknowledgements

This reasch was supported by grants from Department of Sience and Technology of Shandong Province (Grant No. BS2011YY063), and Department of Health of Shandong Province (Grant No. 2011QZ030).

Reference

1) Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA: Cancer J Clin 2010; 60: 277-300.
2) Lerner SP, Skinner DG, Lieskovsky G, Boyd SD, Groshen SL, Ziogas A, Skinner E, Nichols P, Hopwood B. The rationale for en bloc pelvic lymph node dissection for bladder cancer patients with nodal metastases: long-term results. J Urol 1993; 149: 758-764.
3) Ghoneim MA, el-Mekresh MM, el-Baz MA, el-Attar IA, Ashamallah A. Radical cystectomy for carcinoma of the bladder: critical evaluation of the results in 1,026 cases. J Urol 1997; 158: 393-399.
4) Stein JP, Lieskovsky G, Cote R, Groshen S, Feng AC, Boyd S, Skinner E, Bochner B, Thangathurai D, Mikhail M, Raghavan D, Skinner DG. Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. J Clin Oncol 2001; 19: 666-675.
5) Wynder EL, Goldsmith R. The epidemiology of bladder cancer: a second look. Cancer 1977; 40: 1246-1268.
6) Neal DE, Marsh C, Bennett MK, Abel PD, Hall RR, Sainsbury JR, Harris AL. Epidermal-growth-factor receptors in human bladder cancer: comparison of invasive and superficial tumours. Lancet 1985; 325: 366-368.
7) Knowles MA. What we could do now: molecular pathology of bladder cancer. Mol Pathol 2001; 54: 215-221.
8) Dyrskjøt L, Kruhøffer M, Thykjaer T, Marcussen N, Jensen JL, Møller K, Ørntoft TF. Gene Expression in the Urinary Bladder. Cancer Res 2004; 64: 4040-4048.
9) Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A. Human Protein Reference Database--2009 update. Nucleic Acids Res 2009; 37(Database issue): D767-772.
10) Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M. The BioGRID Interaction Database: 2011 update. Nucleic Acids Res 2011; 39(Database issue): D698-704.
11) Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004; 3: Article3.
12) Derrick TR, Bates BT, Dufek JS. Evaluation of time-series data sets using the Pearson product-moment correlation coefficient. Med Sci Sports Exerc 1994; 26: 919-928.
13) Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.
14) Bader GD, Hogue CW. An automated method for finding molecular complexes in largeprotein interaction networks. BMC Bioinformatics 2003; 4: 2.
15) Glatigny A, Mathieu L, Herbert CJ, Dujardin G, Meunier B, Mucchielli-Giorgi MH. An in silico approach combined with in vivo experiments enables the identification of a new protein whose overexpression can compensate for specific respiratory defects in Saccharomyces cerevisiae. BMC Syst Biol 2011; 5: 173.
16) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.
17) Vlahou A, Schellhammer PF, Mendrinos S, Patel K, Kondylis FI, Gong L, Nasim S, Wright GL Jr. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. The Am J Pathol 2001; 158: 1491-1502.
18) Spruck CH 3rd, Ohneseit PF, Gonzalez-Zulueta M, Esrig D, Miyao N, Tsai YC, Lerner SP, Schmütte C, Yang AS, Cote R, Dubeau L, Nichols PW, Hermann GG, Steven K, Horn T, Skinner DG, Jones PA. Two molecular pathways to transitional cell carcinoma of the bladder. Cancer Res 1994; 54: 784-788.
19) Hormozdiari F, Berenbrink P, Przulj N, Sahinalp SC. Not all scale-free networks are born equal: the role of the seed graph in PPI network evolution. PLoS Comput Biol 2007; 3: e118.
20) Jeong H, Mason SP, Barabasi AL, ZN Oltvai ZN. Lethality and centrality in protein networks. Arxiv preprint cond-mat/0105306, 2001.
21) Burge CB, Tuschl T, Sharp PA. 20 Splicing of Precursors to mRNAs by the Spliceosomes. Cold Spring Harbor Monograph Archive 1999; 37: 525-560.
22) Chen M, Von Mikecz A. Formation of nucleoplasmic protein aggregates impairs nuclear function in response to SiO2 nanoparticles. Exp Cell Res 2005; 305: 51-62.
23) Mattaj IW, De Robertis EM. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. Cell 1985; 40: 111-118.
24) Blower MD, Nachury M, Heald R, Weis K. A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. Cell 2005; 121: 223-234.
25) Clancy S. RNA splicing: introns, exons and spliceosome. Nature Education 2008; 1: 1.
26) Will CL, Lührmann R. Spliceosome structure and function. Cold Spring Harbor Perspectives Biology 2011; 3: 3.