European Review for Medical and Pharmacological Sciences 2013; 17: 1044-1050

DNA microarray-based screening of differentially expressed genes related to acute lung injury and functional analysis

Y. CHEN, J.-X. SHI, X.-F. PAN, J. FENG, H. ZHAO

Thoracic Department, Shanghai Chest Hospital, Shanghai, China.
Yong Chen and *Jian-Xin Shi* contribute equally to this work

Abstract. – OBJECTIVES: The purpose of this study was to identify differentially expressed genes (DEGs) related to acute lung injury (ALI) induced by sepsis with DNA microarray.

MATERIALS AND METHODS: Gene expression profile GSE10474 was downloaded from Gene Expression Omnibus (GEO) database which includes 34 samples, among which 13 patients with ALI + sepsis and 21 patients with sepsis alone. The DEGs were identified between ALI + sepsis and sepsis alone samples using R, which were further analyzed using bioinformatics methods. Firstly, HitPredict was used to search protein-protein interactions of the DEGs. Secondly, WebGestalt was adopted for functional enrichment analysis of genes in the interaction networks. Finally, DNA methylation was analyzed to explain the differential expression.

RESULTS: A total of 12 genes were identified as DEGs by comparing chip data from ALI + sepsis samples and those from sepsis alone samples, among which occludin (OCLN) and major histocompatibility complex, class II, DQ beta 1 (HLA-DQB1) had 21 and 6 interactors, respectively. Functional enrichment analysis revealed several significantly over-represented terms: cellular component organization, macromolecular organization and biosynthesis, and response to stimulus. In addition, methylation was found in the promoters of OCLN and HLA-DQB1.

CONCLUSIONS: We screened DEGs in septic ALI samples, and several interesting genes were obtained, especially OCLN and HLA-DQB1. They may be developed into marker genes for diagnosis or treatment of ALI.

Key Words:
Acute lung injury, Sepsis, Bioinformatics, Differentially expressed gene, Interaction network, Functional enrichment analysis, Methylation.

Introduction

Sepsis is a clinical course of systemic inflammatory response caused by bacterial infections, and it is also known as a common complication in critically ill patients after surgery. It usually results in a high morbidity and mortality. Lung is the most vulnerable organ suffering from sepsis, among which acute lung injury (ALI) appears early with a high incidence[1,2]. ALI can be caused by various internal and external pathogenic factors that leading to systemic inflammatory response and it is characterized by refractory hypoxemia and respiratory distress syndrome. Common causes include sepsis, trauma and shock[3], and sepsis is the most common one[4]. Therefore, it is very important to elucidate the pathogenesis of sepsis-induced ALI. Previous study indicates that the pathogenic bacteria ingredients activate a variety of inflammatory cells, which further leads to generations of excessive inflammatory mediators[5]. Another report shows that sepsis induced expression of tissue factor activates coagulation in the lung and leads to a procoagulant environment, which subsequently results in fibrin deposition and inflammatory response. Another work reports that tissue factor complex mainly stimulates the release of inflammatory cytokines and fibrin deposition to promote the animal organ damage[6]. The tissue factor-dependent extrinsic pathway has been suggested as a central mechanism by which the coagulation cascade is locally activated in the lungs of patients with ALI[1]. TNF receptor-associated factor 6 gene, interleukin-6 (IL-6) and IL-1β have also been confirmed to be associated with septic ALI[7]. Though much progress has been achieved in revealing its pathogenesis, more works are needed to describe it systematically.

In terms of treatment, emerging evidence shows that the use of anticoagulants, such as tissue factor pathway inhibitor, antithrombin, heparin, activated protein, and plasminogen activator, especially the tissue-type plasminogen activator, presents a significant effect on improving lung function in ALI and

Corresponding Author: Hang Zhao, MD; e-mail: zhaohangsch@hotmail.com

increased oxygen supply[8]. However, improvements in diagnosis and treatment are still necessary, especially early detection.

Microarray technology is a powerful tool to globally detect the changes in gene expression, which can help to uncover the pathogenesis and advance the discovery of biomarkers. Therefore, chip data from patients with septic ALI were compared with those from patients with sepsis alone to uncover DEGs. Protein-protein interaction analysis and functional enrichment analysis were performed to better characterize their roles in development of ALI.

Materials and Methods

Microarray Data

Data set including 13 samples with ALI + sepsis and 20 samples with sepsis alone was downloaded from NCBI GEO (access no. GSE10474)[9]. The platform is GPL571 [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array. The information about each probe on the chip is provided by Affymetrix, Inc., and gathered along with the raw chip data.

Identification of Differentially Expressed Genes

Raw data was pre-processed by Package Affy[10,11]. And limma[12], an R package, was applied to find out DEGs between ALI + sepsis and sepsis alone. The threshold was set as $p < 0.05$ and $|\log FC| > 1$.

Prediction of Protein-Protein Interactions

Generally, proteins interact with each other to function in different biological processes[13]. Therefore, HitPredict was utilized to identify protein-protein interactions (PPIs) for DEGs. HitPredict combines comprehensive PPIs databases in use, such as IntAct[14], BIOGRID[15] and HPRD[16], containing PPI information from both high-throughput and small-scale experiments. Moreover, HitPredict calculates reliability for each interaction in the form of likelihood ratio using Bayesian algorithm, combining information like sequences, structures and functions[17]. Interaction is regarded of high confidence when the likelihood ratio is bigger than 1.

Functional Enrichment Analysis

Genes in the same network generally interact with each other to achieve the same biological process and function. Since researches focus have transferred from "single genes" to "gene sets", we applied WebGestalt[18,19] that based on the hypergeometric test to perform the enrichment analysis for large sets of differentially expressed gene products.

DNA Methylation Retrieval

DNA methylation is known as a fundamental element of epigenetic modification that is involved in the regulation of gene expressions in the developmental process of multicellular organisms. DNA methylation influences gene activities that circumvent the changes of the DNA sequences[20]. Therefore, DNA methylation was analyzed by searching ENCODE (Encyclopedia of DNA Elements) project in UCSC Genome Browser[21] to elucidate the underlying mechanisms of differentially expressed genes.

Results

Differentially Expressed Genes

R package was used to pre-process the chip data (Figure 1). A total of 12 genes were acquired according to the cut-off ($p < 0.05$ and $|\log FC| > 1$), among which 4 were down-regulated and 8 were up-regulated (Table I).

Interactors of DEGs

HitPredict collects 239,584 PPIs, of which 168,458 are thought to be of high confidence. PPIs of high confidence were retained for the DEGs (Figure 2). Finally, 21 and 6 interactors were identified for OCLN and HLA-DQB1.

Results of Functional Enrichment Analysis for Genes in the Network

For all the genes included in the interaction network, we applied WebGestalt to perform the functional enrichment analysis. Finally, 7 significantly over-represented functions in OCLN network were revealed, which were associated with cell components assembly. And HLA-DQB1 network had 5 significant functions that were related to immune responses (Table II).

DNA Methylation in DEGs

The ENCODE plan uses multiple cell lines to collect all information in the regulatory sequences of the human genome[22]. By USCS Genome Browser query system, sequence analysis of the promoter regions of differentially expressed OCLN and HLA-DQB genes was performed to

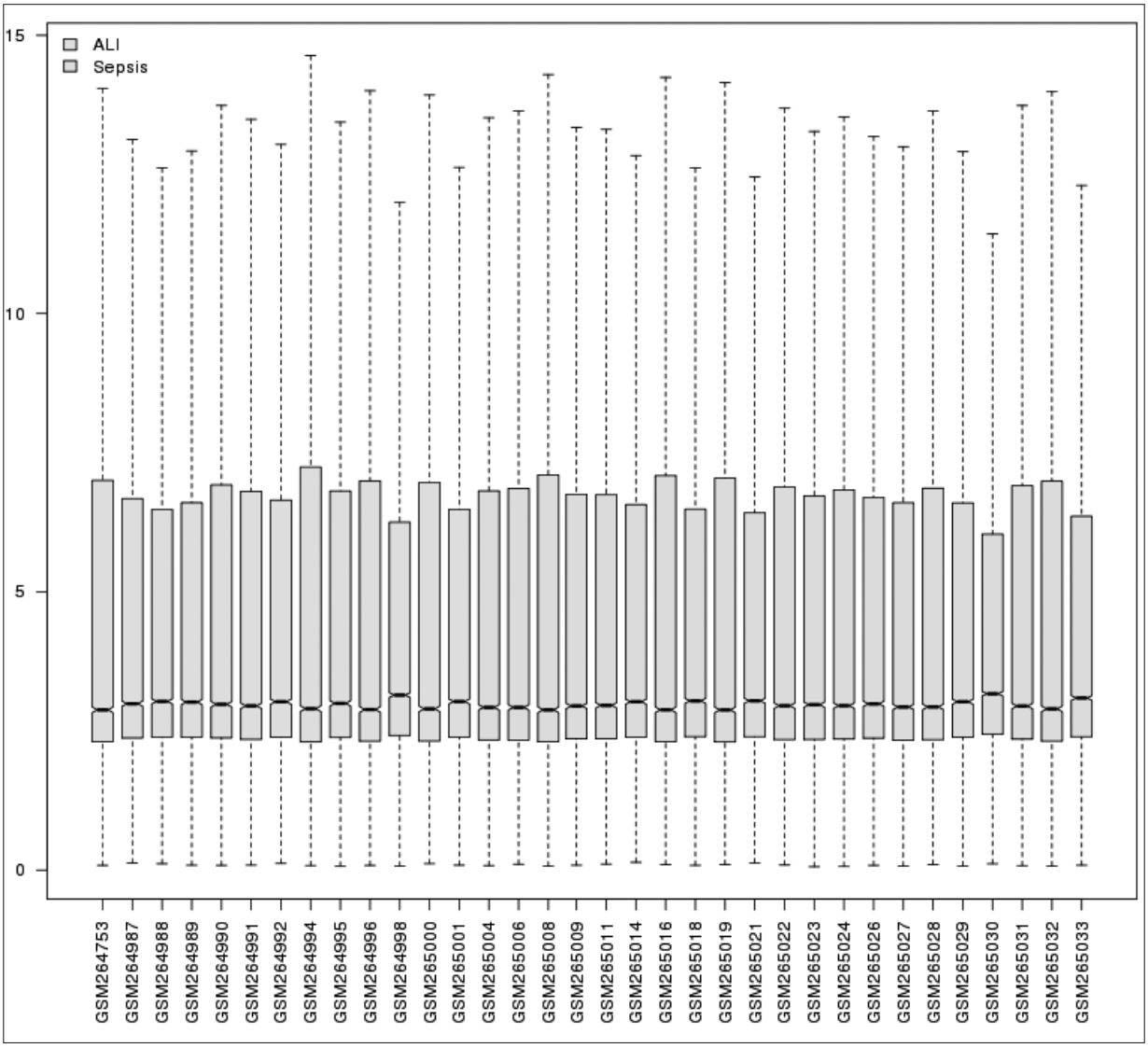

Figure 1. Boxplot of normalized expression data. The blue boxes represent 13 septic ALI samples and the pink boxes for 21 sepsis samples. Black lines in the boxes indicate the medians. They are almost on the same line, suggesting a good performance of the normalization.

acquire information about the status and location of methylation (Figure 3). According to the search results, DNA methylation of *OCLN* CpG island occurs mainly in the promoter region (Figure 3A). As shown in Figure 3B, the methylation region of *HLA-DQB1* gene also locates in the

Table I. List of differentially expressed genes.

Symbol	*p* value	logFC	Symbol	*p* value	logFC
HOPX	0.021837	-1.1871044	TREM1	0.013971	1.08965861
CYBRD1	0.041109	-1.12738059	HIST1H3H	0.038	1.13533773
UPB1	0.01082	-1.0672	CDKN1A	0.000235	1.31803919
OCLN	0.045338	-1.03764029	BTNL8	0.008236	1.46422381
C21orf7	0.034917	1.06017363	HLA-DQB1	0.021759	1.50202784
HIST2H4B	0.036566	1.06655311	CDKN1C	0.004152	1.64387692

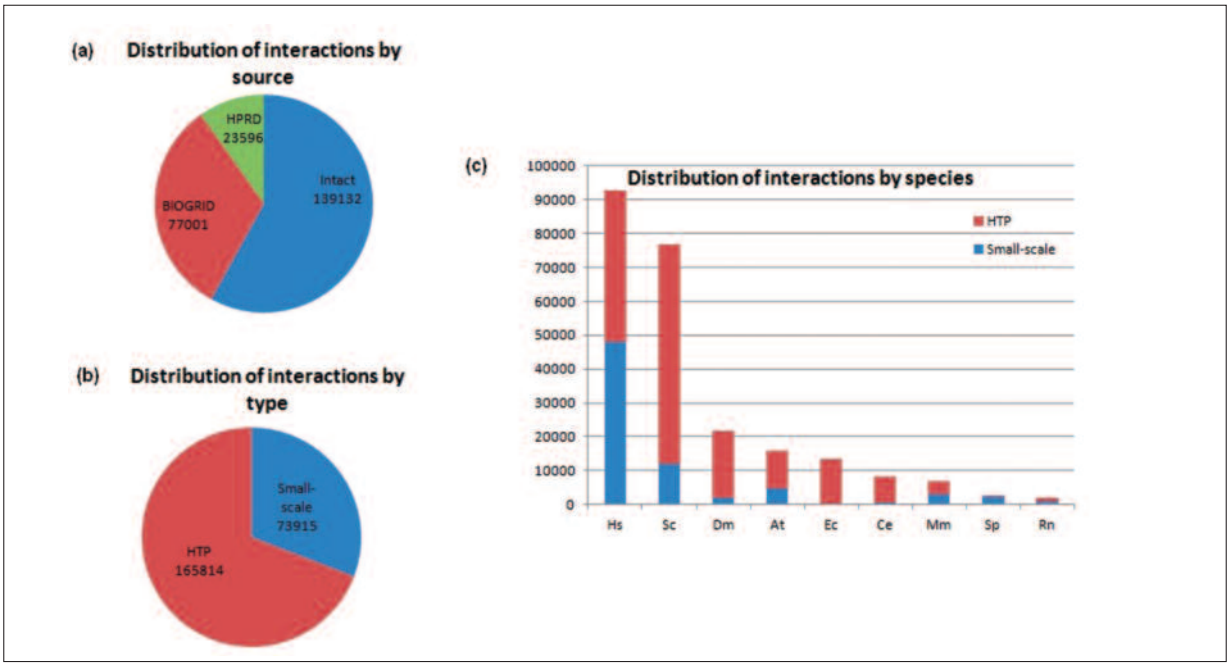

Figure 2. Data sources of PPIs in HitPredict. ***A,*** PPIs from three database, IntAct (blue), BioGRID (*red*) and the HPRD (*green*); ***B,*** PPIs from small-scale (*blue*) or high-throughput (*red*) experiments. ***C,*** PPIs from 9 species in small-scale (*blue*) or high-throughput (*red*) experiments.

promoter region. Given promoter plays an important role in transcriptional regulation, methylation may be a good point for modulation of gene expression and thus treatment of disease.

Discussion

Although significant progress has been made in the treatment of sepsis-induced ALI, disease prevention and timely diagnosis are also important. It is beneficial to find diagnostic marker genes for ALI. In present study, microarray data for patients with septic ALI were compared with that for patients with sepsis alone to screen out genes associated with sepsis-induced ALI. A total of 12 DEGs were obtained, 4 down-regulated and 8 up-regulated (Table I).

Table II. Enrichment analysis in gene interactive network.

Function	*p*	adjP
OCLN		
Cellular component organization	rawP = 0.0009	adjP = 0.0151
Protein complex assembly	rawP = 0.0009	adjP = 0.0151
Protein complex biogenesis	rawP = 0.0009	adjP = 0.0151
Macromolecular complex assembly	rawP = 0.0020	adjP = 0.0180
Macromolecular complex subunit organization	awP = 0.0027	adjP = 0.0182
Cellular component assembly	rawP = 0.0046	adjP = 0.0226
Cellular component biogenesis	rawP = 0.0064	adjP = 0.0258
HLA-DQB1		
Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	rawP = 1.96e-08	adjP = 3.14e-06
Immune response	rawP = 3.82e-07	adjP = 3.06e-05
Antigen processing and presentation	rawP = 7.23e-07	adjP = 3.86e-05
Immune system process	rawP = 2.23e-06	adjP = 8.92e-05
Response to stimulus	rawP = 0.0008	adjP = 0.0033

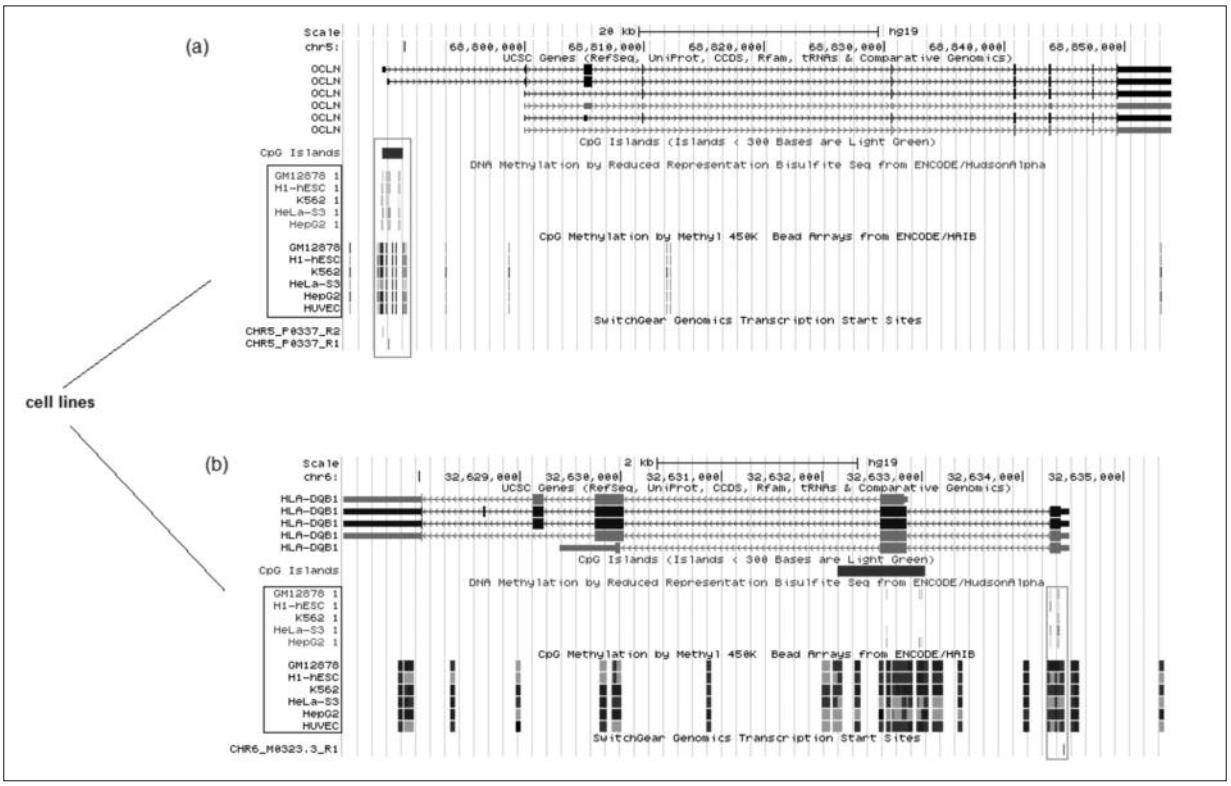

Figure 3. Methylation sites in OCLN and HLA-DQB1. The black box contains information of cell lines used in the ENCODE project. The methylation sites are indicated by red rectangles on the chromosome diagram. ***A,*** OCLN. ***B,*** HLA-DQB.

OCLN is an integral membrane protein that is required for cytokine-induced regulation of tight junction, and its down-regulation is indicative of lung injury. You et al[23] find that OCLN messenger RNA and protein expressions in lung tissue are consistent with the degree of lung injury. The study by Shimada et al[24] reports that OCLN is down-regulated in diffuse alveolar damage induced by hyperoxia exposure in mice. TREM1 is a receptor involved in amplification of neutrophil and monocytes-mediated inflammatory response[25]. It was found to be up-regulated in present study in septic ALI, which was in accordance with previous findings[26]. It has been thought to be a potential therapeutic target in diseases characterized by an excessive inflammatory response[25]. CDKN1A (p21) is a regulator of cell cycle progression at G1. Its expression is controlled by p53, through which this protein mediates the p53-dependent cell cycle G1 phase arrest in response to a variety of stress stimuli[27]. It also takes a part in DNA replication and DNA damage repair[28]. Therefore, it's not difficult to comprehend its up-regulation in ALI where excessive inflammatory response occurs, as well as CDKN1C (p57). HLA-DQB1 (major histocompatibility complex, class II, DQ beta 1) plays a critical role in the immune system by presenting peptides derived from extracellular proteins[29], so the up-regulation is also expectable. Besides, two histones were also identified as up-regulated genes: HIST2H4B and HIST1H3H. This may explain the global changes in gene expression as they are responsible for nucleosome structure of the chromosomal fiber and subsequently affect the transcription.

To further uncover their roles in the development of septic ALI, interactors of the DEGs were retrieved by HitPredict. Finally, 21 interactors were acquired for OCLN and 6 for HLA-DQB1, and corresponding interaction networks were also established. Functional enrichment analysis was carried out for these genes with WebGestalt. 7 significantly over-represented terms were revealed for OCLN, and they were associated with cellular component organization and protein complex assembly. Some of the interactors for OCLN are implicated in cell junction and cell communication. Gap junction protein, beta 1 (GJB1) is a member of the gap junction protein

family. Previous reports have indicated that the regulation of gene expression of gap junction proteins is affected during injury[30]. SMAD family member 4 (SMAD4) is a member of the Smad family of signal transduction proteins, which are phosphorylated and activated by transmembrane serine-threonine receptor kinases in response to TGF-beta signaling[31,32]. Epidermal growth factor receptor (EGFR) pathway substrate 15 (EPS15) participates in receptor-mediated endocytosis of EGF[33]. Epsin 1 (EPN1) binds clathrin and is involved in the endocytosis of clathrin-coated vesicles[34]. 5 significant terms were disclosed for HLA-DQB1, all of which were associated with immune response. Several interactors were associated with class II major histocompatibility complex, such as major histocompatibility complex, class II, DQ alpha 2 (HLA-DQA2), CD74 (an important chaperone that regulates antigen presentation for immune response[35]). CD4 was also included in the list, which initiates the early phase of T-cell activation. According to the functions of some typical interactors, it could be speculated that OCLN is indicative of lung injury and HLA-DQB1 can reflect the degree of immune response. Thus both of them are worthy of further researches.

In addition, DNA methylation was investigated to explain the differential expression of OCLN and HLA-DQA2. Methylation was found in promoters for both genes, which might be developed into targets for modulation of the expression. Previous studies have reported the methylation of OCLN in cancers[36,37]. Methylation of HLA-DQA2 is also investigated[38,39]. More studies are beneficial to determine their roles in development ALI.

Conclusions

Overall, chip data from patients with septic ALI were compared with those from patients with sepsis alone to screen out DEGs. Protein-protein interaction analysis and functional enrichment analysis were performed and two interesting DEGs were revealed: OCLN and HLA-DQA2. OCLN may be used for diagnosis and monitoring of ALI while HLA-DQA2 is potential drug target to modulate immune response. Besides, DNA methylation may contribute to the differential expression of the two genes, and more studies are necessary to deepen the understanding of the regulation mechanism.

References

1) Morris PE, Steingrub JS, Huang BY, Tang S, Liu PM, Rhode PR, Wong HC. A phase I study evaluating the pharmacokinetics, safety and tolerability of an antibody-based tissue factor antagonist in subjects with acute lung injury or acute respiratory distress syndrome. BMC Pulm Med 2012; 12: 5.
2) Maybauer MO, Traber DL, Maybauer DM. Recombinant human activated protein C in acute lung injury: what is the role of bronchial circulation? Crit Care 2009; 13: 112.
3) Barbas CS. Acute lung injury and acute respiratory distress syndrome: diagnostic hurdles. J Bras Pneumol 2007; 33: xxv-xxvi.
4) Atabai K, Matthay MA. The pulmonary physician in critical care. 5: Acute lung injury and the acute respiratory distress syndrome: definitions and epidemiology. Thorax 2002; 57: 452-458.
5) Liaudet L, Murthy KG, Mabley JG, Pacher P, Soriano FG, Salzman AL, Szabo C. Comparison of inflammation, organ damage, and oxidant stress induced by Salmonella enterica serovar Muenchen flagellin and serovar Enteritidis lipopolysaccharide. Infect Immun 2002; 70: 192-198.
6) Welty-Wolf KE, Carraway MS, Miller DL, Ortel TL, Ezban M, Ghio AJ, Idell S, Piantadosi CA. Coagulation blockade prevents sepsis-induced respiratory and renal failure in baboons. Am J Respir Crit Care Med 2001; 164(10 Pt 1): 1988-1996.
7) Zhao D, Ding R, Mao Y, Wang L, Zhang Z, Ma X. Heparin rescues sepsis-associated acute lung injury and lethality through the suppression of inflammatory responses. Inflammation 2012; 35: 1825-1832.
8) MacLaren R, Stringer KA. Emerging role of anticoagulants and fibrinolytics in the treatment of acute respiratory distress syndrome. Pharmacotherapy 2007; 27: 860-873.
9) Howrylak JA, Dolinay T, Lucht L, Wang Z, Christiani DC, Sethi JM, Xing EP, Donahoe MP, Choi AM. Discovery of the gene signature for acute lung injury in patients with sepsis. Physiol Genomics 2009; 37: 133-139.
10) Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. Bioinformatics 2001; 17: 520-525.
11) Fujita A, Sato JR, Rodrigues Lde O, Ferreira CE, Sogayar MC. Evaluating different methods of microarray data normalization. BMC Bioinformatics 2006; 7: 469.
12) Smyth GK. Limma: linear models for microarray data. In: Gentleman R, et al Eds. Bioinformatics and computational biology solutions using R and bioconductor. R. Springer, New York, 2005; pp. 397-420.
13) Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A, Legrain P. The protein-protein interaction map of Helicobacter pylori. Nature 2001; 409(6817): 211-215.

14) Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeiffenberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H. The IntAct molecular interaction database in 2012. Nucleic Acids Res 2012; 40(Database issue): D841-846.

15) 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.

16) Mathivanan S, Ahmed M, Ahn NG, Alexandre H, Amanchy R, Andrews PC, Bader JS, Balgley BM, Bantscheff M, Bennett KL, Bjorling E, Blagoev B, Bose R, Brahmachari SK, Burlingame AS, Bustelo XR, Cagney G, Cantin GT, Cardasis HL, Celis JE, Chaerkady R, Chu F, Cole PA, Costello CE, Cotter RJ, Crockett D, DeLany JP, De Marzo AM, DeSouza LV, Deutsch EW, Dransfield E, Drewes G, Droit A, Dunn MJ, Elenitoba-Johnson K, Ewing RM, Van Eyk J, Faca V, Falkner J, Fang X, Fenselau C, Figeys D, Gagne P, Gelfi C, Gevaert K, Gimble JM, Gnad F, Goel R, Gromov P, Hanash SM, Hancock WS, Harsha HC, Hart G, Hays F, He F, Hebbar P, et al. Human Proteinpedia enables sharing of human protein data. Nat Biotechnol 2008; 26: 64-167.

17) Patil A, Nakai K, Nakamura H. HitPredict: a database of quality assessed protein-protein interactions in nine species. Nucleic Acids Res 2011; 39(Database issue): D744-749.

18) Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 2005; 33(Web Server issue): W741-748.

19) Duncan D, Prodduturi N, Zhang B. WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. BMC Bioinformatics 2010; 11(Suppl 4): P10.

20) Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell 2007; 128: 635-638.

21) Boyle AP, Song L, Lee BK, London D, Keefe D, Birney E, Iyer VR, Crawford GE, Furey TS. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res 2011; 21: 456-464.

22) Pennisi E. Genomics. ENCODE project writes eulogy for junk DNA. Science 2012; 337(6099): 1159, 1161.

23) You K, Xu X, Fu J, Xu S, Yue X, Yu Z, Xue X. Hyperoxia disrupts pulmonary epithelial barrier in newborn rats via the deterioration of occludin and ZO-1. Respir Res 2012; 13: 36.

24) Shimada I, Matsui K, Brinkmann B, Hohoff C, Hiraga K, Tabuchi Y, Takasaki I, Kato I, Kawaguchi H, Takasawa K, Iida R, Takizawa H, Matsuki T. Novel transcript profiling of diffuse alveolar damage induced by hyperoxia exposure in mice: normalization by glyceraldehyde 3-phosphate dehydrogenase. Int J Legal Med 2008; 122: 373-383.

25) Colonna M, Facchetti F. TREM-1 (triggering receptor expressed on myeloid cells): a new player in acute inflammatory responses. J Infect Dis 2003; 187(Suppl 2): S397-401.

26) Sadikot RT, Christman JW, Blackwell TS. Molecular targets for modulating lung inflammation and injury. Curr Drug Targets 2004; 5: 581-588.

27) Fournier C, Wiese C, Taucher-Scholz G. Accumulation of the cell cycle regulators TP53 and CDKN1A (p21) in human fibroblasts after exposure to low- and high-LET radiation. Radiat Res 2004; 161: 675-684.

28) Cazzalini O, Scovassi AI, Savio M, Stivala LA, Prosperi E. Multiple roles of the cell cycle inhibitor p21(CDKN1A) in the DNA damage response. Mutat Res 2010; 704: 12-20.

29) Kwok WW, Domeier ME, Johnson ML, Nepom GT, Koelle DM. HLA-DQB1 codon 57 is critical for peptide binding and recognition. J Exp Med 1996; 183: 1253-1258.

30) De Maio A, Vega VL, Contreras JE. Gap junctions, homeostasis, and injury. J Cell Physiol 2002; 191: 269-282.

31) Itoh S, ten Dijke P. Negative regulation of TGF-beta receptor/Smad signal transduction. Curr Opin Cell Biol 2007; 19: 176-184.

32) Moustakas A, Souchelnytskyi S, Heldin CH. Smad regulation in TGF-beta signal transduction. J Cell Sci 2001; 114(Pt 24): 4359-4369.

33) Salcini AE, Chen H, Iannolo G, De Camilli P, Di Fiore PP. Epidermal growth factor pathway substrate 15, Eps15. Int J Biochem Cell Biol 1999; 31: 805-809.

34) Horvath CA, Vanden Broeck D, Boulet GA, Bogers J, De Wolf MJ. Epsin: inducing membrane curvature. Int J Biochem Cell Biol 2007; 39: 1765-1770.

35) Beswick EJ, Reyes VE. CD74 in antigen presentation, inflammation, and cancers of the gastrointestinal tract. World J Gastroenterol 2009; 15: 2855-2861.

36) Ding S, Gong BD, Yu J, Gu J, Zhang HY, Shang ZB, Fei Q, Wang P, Zhu JD. Methylation profile of the promoter CpG islands of 14 "drug-resistance" genes in hepatocellular carcinoma. World J Gastroenterol 2004; 10: 3433-3440.

37) Yu YP, Paranjpe S, Nelson J, Finkelstein S, Ren B, Kokkinakis D, Michalopoulos G, Luo JH. High throughput screening of methylation status of genes in prostate cancer using an oligonucleotide methylation array. Carcinogenesis 2005; 26: 471-479.

38) Majumder P, Boss JM. DNA methylation dysregulates and silences the HLA-DQ locus by altering chromatin architecture. Genes Immun 2011; 12: 291-299.

39) Hernandez-Vargas H, Lambert MP, Le Calvez-Kelm F, Gouysse G, McKay-Chopin S, Tavtigian SV, Scoazec JY, Herceg Z. Hepatocellular carcinoma displays distinct DNA methylation signatures with potential as clinical predictors. PLoS One 2010; 5: e9749.