Original Article
Identification of candidate genes and pathways in cisplatin-resistant NSCLC by integrated bioinformatics analysis

Fang Yu[1*], Chengtao Fan[2*], Yinghong Ren[1]

Departments of [1]Internal Medicine, [2]Cardiothoracic Surgery, Shangluo Central Hospital, Shangluo, Shaanxi, China. *Equal contributors.

Abstract: Cisplatin resistance is a main barrier in the treatment of non-small cell lung cancer (NSCLC), but the underlying mechanisms remain poorly understood. This study analyzed transcriptome data in cisplatin-resistant NSCLC cells and identified differentially expressed genes (DEGs), which were further subjected to gene ontology (GO) and pathway enrichment analysis, protein-protein interaction network establishment and survival analysis. A total of 235 DEGs (99 up-regulated and 136 down-regulated) were identified from the GSE21656 dataset. PPI network identified 9 hub genes correlated significantly with patients' prognosis. In conclusion, this study identified candidate genes and pathways associated with the development of cisplatin resistance. These findings provide potential targets for improving efficacy of cisplatin-based chemotherapy in NSCLC patients.

Keywords: Lung cancer, cisplatin resistance, bioinformatics analysis

Introduction

Lung cancer, with high incidence and mortality, is the leading cause of cancer deaths worldwide. Each year, more than 2.2 million patients are diagnosed with lung cancer [1]. Non-small cell lung cancer (NSCLC) is the most common type of lung cancer which has an estimated 5-year survival rate of ~10% [2, 3]. Cisplatin-based chemotherapy improves the prognosis of NSCLC, but the efficacy is limited due to the development of cisplatin resistance [4]. However, the underlying mechanism of this resistance remains unclear.

Microarray represents a high throughput analysis platform which enables investigators to measure the expression levels of large numbers of genes simultaneously and has been widely used in the field of life science [5]. With the extensive application of microarray, a number of studies have explored the gene expression signature of cisplatin resistant lung cancer cells and have revealed hundreds of genes that were differentially expressed. To gain compre-

hensive understanding of the differentially expressed genes (DEGs), bioinformatics analysis, which can provide insights into the potentially involved biological processes and pathways [6] has been often applied.

In the current study, the original microarray dataset GSE21656 [7] from The Gene Expression Omnibus database (GEO, http://www.ncbi.nlm.nih.gov/geo) was downloaded. Gene expression profiles of cisplatin resistant lung cancer cells were compared with non-resistant lung cancer cells; DEGs were obtained using GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/). Bioinformatics analysis and survival analysis were then performed to identify hub genes involved in the development of cisplatin resistance.

Materials and methods

Microarray data

The gene expression profiles of GSE21656 were downloaded from the GEO database. GSE21656 was based on GPL6244 platform
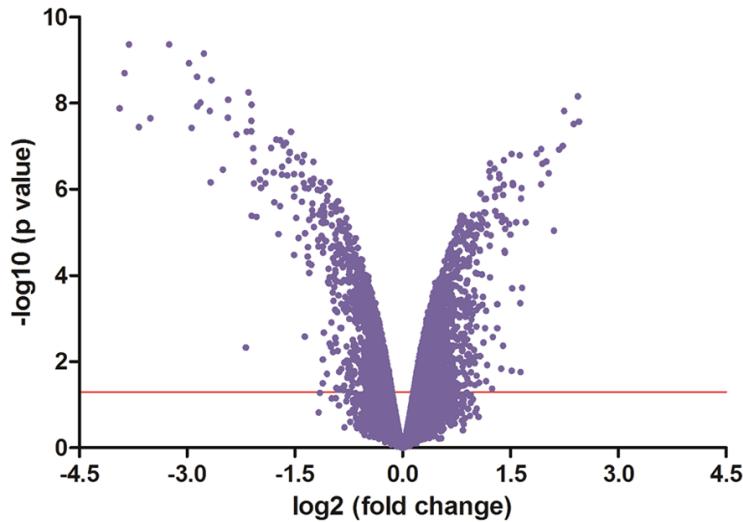
Figure 1. Volcano plot of cisplatin resistance associated genes. Log2 fold change and negative (-) log10 *p*-value were plotted on the x- and y-axis, respectively. Each gene is represented by a dot. The horizontal line indicates P=0.05.

(Affymetrix Human Gene 1.0 ST Array) and the dataset contains 3 biological replicates of each cell line (cisplatin resistant lung cancer cells and non-resistant cancer cells).

Data processing

To identify DEGs between cisplatin-resistant and non-resistant samples, GEO2R, an interactive web tool that allows the identification of genes that are differentially expressed across experimental conditions with *Linear Models for Microarray Data (limma)* was used. P<0.05 and [logFC] >1 were applied to define significant DEGs.

Gene ontology and pathway enrichment analysis

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of candidate DEGs were performed using an internet tool suite-The Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/), which provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes.

Protein-protein interaction (PPI) network establishment, module analysis, and significant candidate genes identification.

PPI network of DEGs in the present study was constructed using NetworkAnalyst (http://www.networkanalyst.ca), a webbased tool that allows investigators to easily navigate large complex gene expression data sets to identify important connections, patterns and functions. First-order interaction network (this network type composes only the seed nodes and the nodes that interact directly with them) was applied and module analysis was then performed to determine the key modules and hub genes.

Expression validation and survival analysis of hub DEGs

To assess expression and prognostic value of the hub genes in the PPI network, GEPIA (http://gepia.cancer-pku.cn/) and Kaplan-Meier plotter (http://kmplot.com/analysis/) were used. The hazard ratio (HR) with 95% confidence intervals were calculated and log rank test was employed to compare the survival distribution of two groups (high and low expression of a gene of interest).

Results

Identification of DEGs

The dataset GSE21656 was analyzed with GEO2R and the expression level of each gene transformed with log2-fold change was calculated (**Figure 1**), the top 250 genes ranked by *p* value were selected for further evaluation. After removal of duplicates and genes that were unidentified, a total of 235 DEGs (99 up-regulated and 136 down-regulated) were finally included (**Table 1**).

Signaling pathway and gene ontology enrichment analysis

To advance understanding of cisplatin resistance, the identified DEGs were uploaded to the online program DAVID. The results show that the up-regulated DEGs were significantly enriched in arrhythmogenic right ventricular cardiomyopathy, cell adhesion molecules, and

Table 1. List of the 235 differentially expressed genes (DEGs)

| Expression	Differentially Expressed Genes
Up-regulated (n=99)	GJA5, CNTN1, COL3A1, ATP11A, ANTXR2, ADARB1, PTPN20, ROBO1, PHLDB2, TPM2, PHACTR2, CD302, FAM83A, CP, SLPI, IGF2BP1, TGM2, CHRNA9, ADGRG3, SNAP25, MYH10, IGFBP4, BMP5, ANO5, ACP6, TENM3, NLGN4Y, KCNK1, COL12A1, TGFBR3, KIAA1324L, SCUBE1, GFPT2, VAV3, MYO1E, TXLNGY, SLC26A2, PDK4, CA2, FLRT3, CDK14, TMTC1, SUCNR1, TM4SF19, PKP2, PTPRN, CALB2, PAPPA2, TIPARP, RASGRF2, DSP, FGD6, ITGB8, CPLX2, MRAP2, PKIA, CDH2, ALX1, ERO1B, AADAC, E2F7, SPINK6, RSPO3, EWSAT1, LRP1B, RALGPS2, NEO1, SPINK5, SLC16A5, IL1A, LPCAT2, SLC2A1, HIST1H2BM, CSRNP3, HEXA, LGSN, USP9Y, ALDH2, RNF217, SH3RF1, VWDE, RPS4Y1, LAMB3, SLC25A15, RASAL2, PLEKHA2, ADGRG6, DDX3Y, TMBIM4, CHMP4C, CSNK1E, SLC22A5, KCNQ5, CKMT1A, RHOF, GYG2, FBN2, OSBPL10, GREB1L
Down-regulated (n=136)	SLFN5, GLRX, ALCAM, HES7, CNTNAP3B, NR4A3, CTH, TGFBR1, HSPB6, CALM1, CA12, SERPINB8, TFPI2, TM4SF1, AKR1C1, CTSL, CHD1, CDH6, TMX4, PIK3R1, RNF212, PLCB1, SRPX, SLC4A11, CHN2, IGFBP2, ANTXR1, DUSP6, PDE3A, DUSP1, RBPMS, KCNE4, C16orf62, CBR1, L1CAM, VEGFC, RARB, NT5E, DHRS3, SPATA7, LXN, IFITM2, AQP3, TNFRSF19, RGL1, CFH, TSPAN13, FRMD6, TMOD1, PAG1, NOTCH3, SEMA3D, NAP1L3, ANO3, NKAIN2, TEX15, PRKCB, AKR1C2, SGCE, SEMA5A, CPVL, SOX6, RAB39B, DKK1, MLPH, SERPINB2, IL1R1, APLN, ADAMTS9, FSTL1, PLAT, ADGRD1, MYO5C, DCLK1, UGT8, DOC2B, SLCO5A1, PRICKLE1, WARS, CRYAB, TC2N, SERPIND1, HPGDS, FGF14, FRY, GPR65, AMDHD1, TUBB4A, PLAC4, SLC16A6, CDH11, GALC, ESM1, FAM107B, S100A16, AKR1C3, ELMO1, NAV2, CACNA2D3, NCAM2, NEDD9, NPR3, FADS2, ABCA8, CDR1, ADGRL2, MACC1, IL18R1, ZEB2, INPP4B, ZNF804A, SLC38A5, PPEF1, IGFBP3, PEG10, LIN7A, PAPPA, CD24, EPHA4, FST, SH2D2A, KCNK2, ADGRV1, CHL1, SCN3A, MCAM, TMEM27, TMPRSS15, ALDH1A1, ACSS3, PLA2G7, RAB3C, FAM216B, SLFN11, SERPINB11, NTS
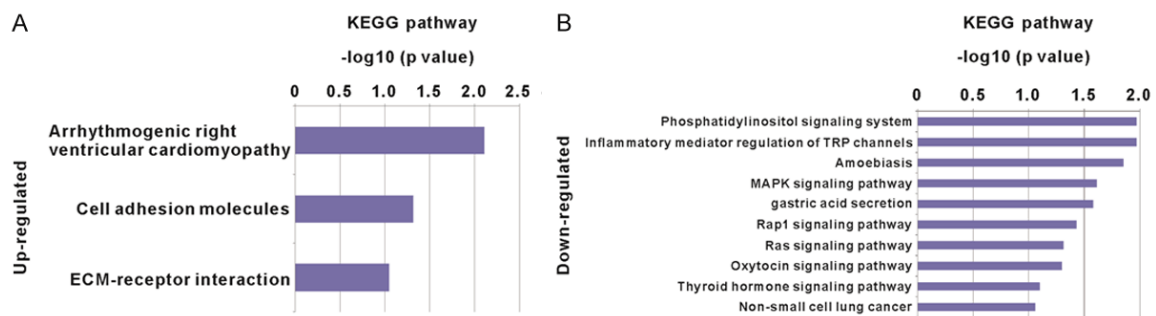
Figure 2. KEGG pathway analysis of the candidate DEGs. Significant enriched pathways of up-regulated (A) and down-regulated DEGs (B).

ECM-receptor interaction (**Figure 2A**), while the most significantly enriched pathways of the down-regulated DEGs were phosphatidylinositol signaling system, inflammatory mediator regulation of TRP channels, amoebiasis, MAPK signaling pathway, gastric acid secretion, and Rap1 signaling pathway, etc (**Figure 2B**). GO analysis demonstrated that up-regulated DEGs were significantly enriched in biological processes including cell adhesion and extracellular matrix degradation (**Figure 3A**); cellular components including plasma membrane, extracellular exosome, and extracellular matrix (**Figure 3B**); molecular functions including guanyl-nucleotide exchange factor activity, cell adhesive protein binding, cell adhesion molecule binding (**Figure 3C**). For down-regulated DEGs, cell adhesion, cellular response to jasmonic acid stimulus, and signal transduction were the most significantly enriched biological

processes (**Figure 3D**); extracellular exosome, cell surface, and plasma membrane were the most significantly enriched cellular components (**Figure 3E**); oxidoreductase activity, ketosteroid monooxygenase activity, and dehydrogenase activity were the most significantly enriched molecular functions (**Figure 3F**).

PPI network establishment and module analysis

The PPI network, constructed with NetworkAnalyst, containing 719 nodes and 821 edges and was further subjected to module detection. The most important modules of DEGs are highlighted in **Figure 4**. Genes within a module were more likely to work together in performing a biological function. The hub genes of each module were considered to have a high hierarchical role in the network. The up-regulated hub genes
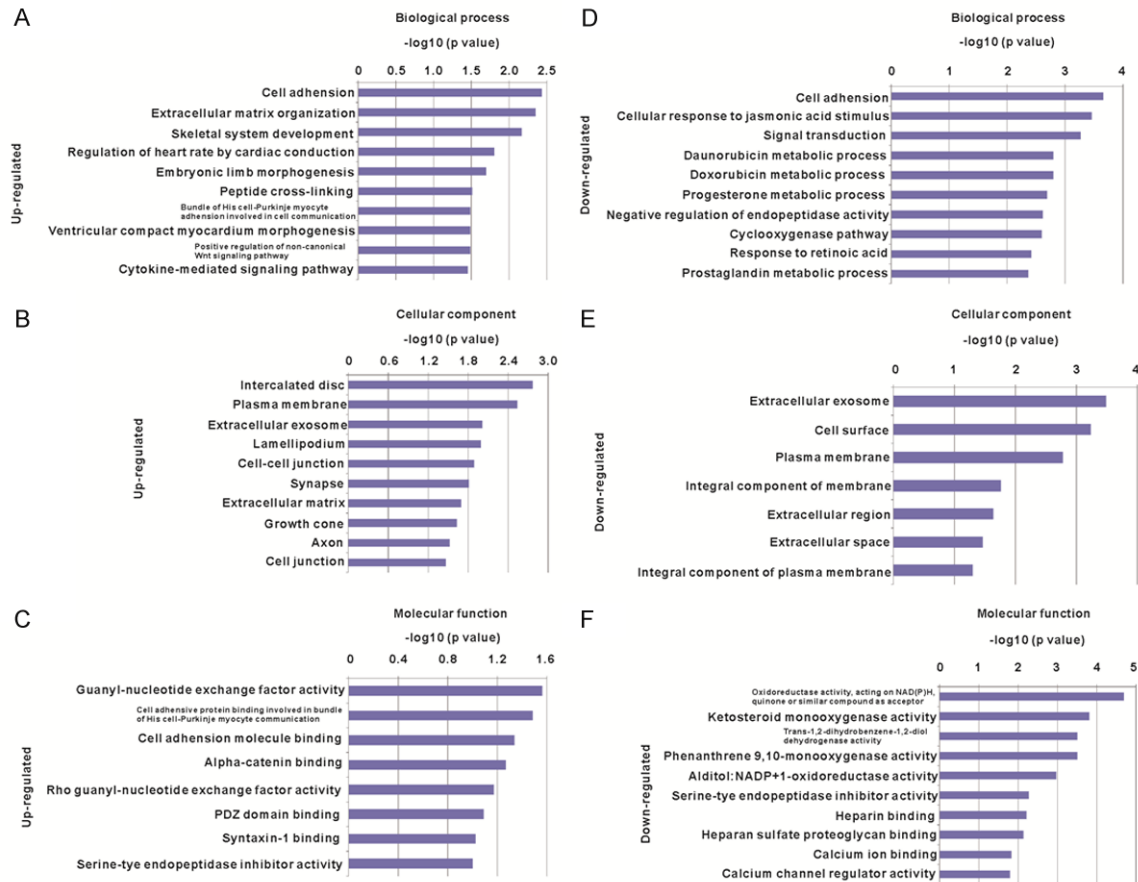
Figure 3. GO analysis of the candidate DEGs. Enriched biological process, cellular component and molecular function of up-regulated (A-C) and down-regulated (D-F) DEGs.

included cadherin 2 (CDH2), casein kinase 1 epsilon (CSNK1E), tropomyosin 2 (TPM2), ras homolog family member F (RHOF), and ribosomal protein S4, Y-linked 1 (RPS4Y1), while the down-regulated hub genes included protein kinase C beta (PRKCB), transforming growth factor beta receptor 1 (TGFBR1), phospholipase C beta 1 (PLCB1), calmodulin 1 (CALM1), NOTCH3, and tubulin beta 4A class IVa (TUBB4A).

Expression validation and survival analysis

To validate expression levels of the 11 hub genes, the TCGA-based web server-GEPIA was used. As shown in **Figure 5**, expression of NOTCH3 and PLCB1 were significantly up-regulated in lung squamous cell carcinoma (LUSC), while expression of PRKCB and RHOF were down-regulated in LUSC. To gain insight into the prognostic value of the 4 hub genes, the online

program Kaplan-Meier plotter was used to evaluate the overall survival of patients with NSCLC according to the expression level (high or low) of each gene. The results show that high mRNA expression of PLCB1 (HR=0.59 [0.47-0.75], P=1.1e-5) and PRKCB (HR=0.54 [0.43-0.69], P=4.1e-7) are associated with better overall survival for NSCLC patients, while increased mRNA expression of NOTCH3 (HR=2.07 [1.63-2.63], P=1.2e-9) and RHOF (HR=1.88 [1.48-2.39], P=1.4e-7) are unfavorable prognostic predictors of overall survival for patients with NSCLC (**Figure 6**).

Discussion

NSCLC accounts for about 85% of total lung cancer cases with a low 5-year survival rate [8]. Cisplatin, by inducing cancer cell apoptosis, represents a conventional treatment for NSCLC [9, 10]. However, its efficacy is weakened due
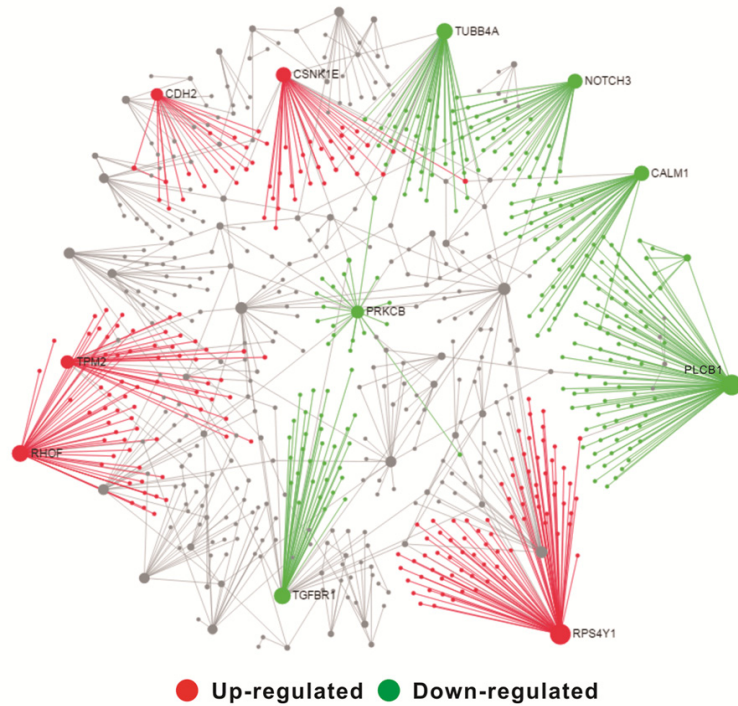
Figure 4. PPI network of DEGs. Using the NetworkAnalyst online program, a topology-based PPI network was established. Hub genes were identified by module detection and highlighted in dark gray (up-regulated) or light gray (down-regulated).

enriched in cell adhesion and cellular response to jasmonic acid stimulus. Adhesion is a hallmark of solid cancer cells and enables tumor growth, survival, and metastasis. It is known that cell adhesion can mediate drug resistance in various kinds of cancers, including NSCLC [11]. MAPK signaling pathway governs cell proliferation, differentiation, cell death, and survival and therefore represents a potential therapeutic target in NSCLC [12].

To understand the interaction between the DEGs, the PPI network was constructed and module detection was performed. These analyses revealed 11 hub genes: CDH2, TPM2, CSNK1E, RHOF, RPS4Y1, TUBB4A, NOTCH3, CALM1, PLCB1, TGFBR1, and PRKCB. Expression validation revealed that NOTCH3 and PLCB1 were significantly up-regulated in LUSC, while expression of PRKCB and RHOF were down-regulated in LUSC. Further survival analysis showed that high mRNA expression of NOTCH3 and RHOF were associated with a poor clinical prognosis, while increased expression of PLCB1 and PRKCB indicates a favorable prognosis. Enhanced expression of protein encoded by NOTCH3 has been implicated in tumor growth and metastasis, and even chemoresistance [13-15]. RHOF is a member of the Rho-GTPase family and higher levels of RHOF have been detected in neoplastic cells and tissues of B-cell origin [16]. PLCB1 plays an important role in the intracellular transduction of extracellular signals and can be regulated by early growth response transcription factor EGR1 [17]. PRKCB is a member of the protein kinase C family and members in this family are known to be involved in diverse cellular signaling pathways. Du et al. reported that PRKCB is critical for bladder cancer cell invasion and migration [18], while in pancreatic cancer, PRKCB was found act as an important suppressor of tumorigenic behavior [19].

to the development of cisplatin resistance and the underlying mechanism remains poorly understood. Microarray can provide expression levels of large numbers of genes simultaneously and has been broadly used in various diseases to reveal the underlying mechanisms, identify the potential therapeutic targets and prognostic markers [5].

In the current study, 99 up-regulated DEGs and 136 down-regulated DEGs were identified to be associated with the development of cisplatin resistance from GSE21656. KEGG pathway analysis demonstrated that arrhythmogenic right ventricular cardiomyopathy, cell adhesion molecules, and ECM-receptor interaction were the enriched pathways of up-regulated DEGs, while down-regulated genes were mainly enriched in the phosphatidylinositol signaling system, inflammatory mediator regulation of TRP channels, amoebiasis, and MAPK signaling pathway. In addition, functional analysis revealed that up-regulated DEGs were chiefly involved in cell adhesion and extracellular matrix organization, while down-regulated DEGs were mainly
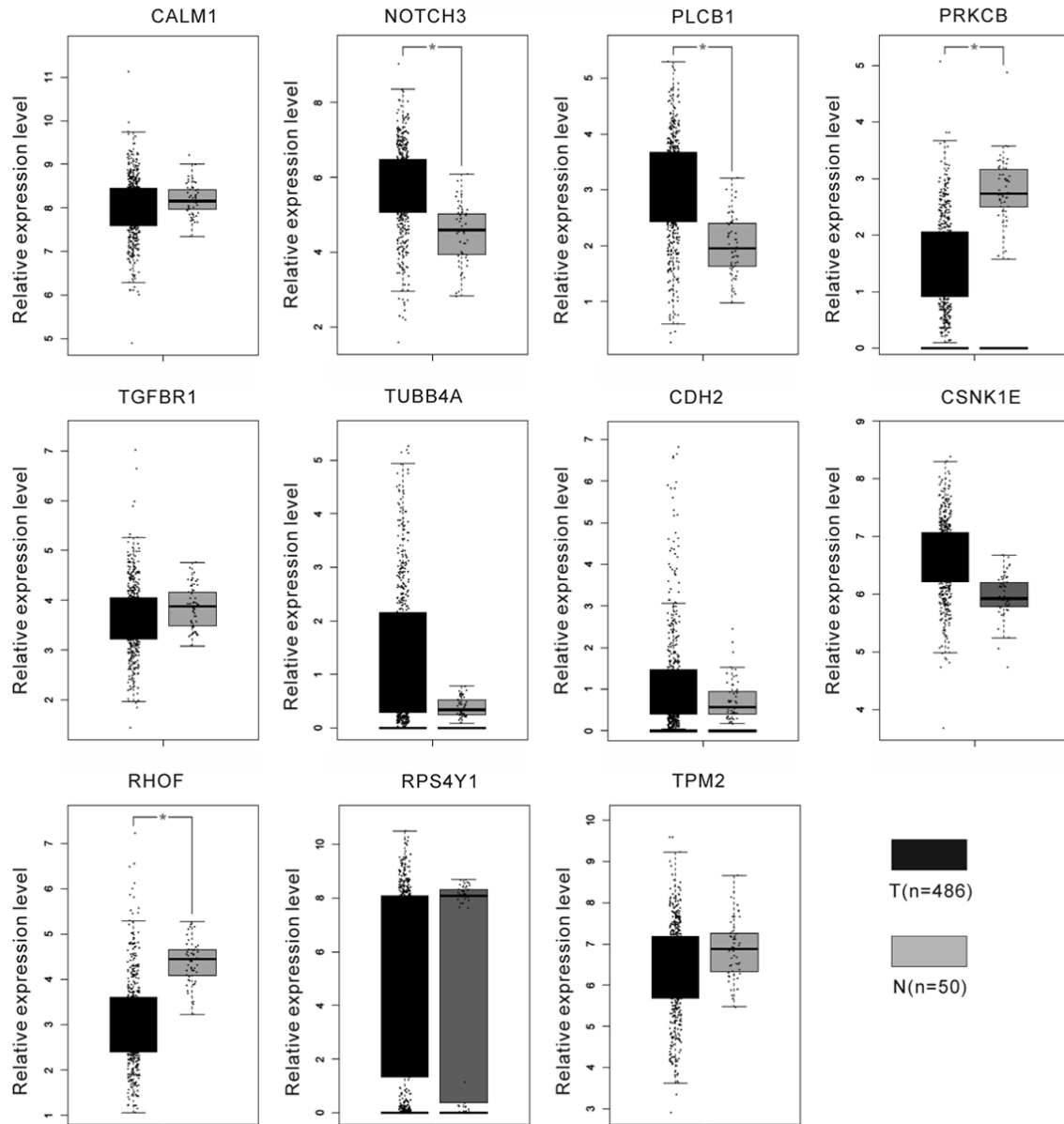
Figure 5. Expression validation. Expression levels of the 11 hub genes were obtained from GEPIA and presented as boxplots.

In conclusion, combining gene expression data and bioinformatics analysis, 235 DEGs were identified as candidate genes that may be implicated in the development of cisplatin resistance. Protein-protein interaction network and functional module detection revealed 11 DEGs as hub genes, among which 4 genes were significant correlated with the survival of NSCLC patients. These findings may advance our understanding of the molecular events underlying cisplatin resistance and provide sets of promising targets for improving lung cancer response to cisplatin based therapy. Future investigations, however, are needed to confirm the function of these genes and promote clinical translation.

Disclosure of conflict of interest

None.

Address correspondence to: Yinghong Ren, Department of Internal Medicine, Shangluo Central Hospital, 148 Beixin Street, Shangluo 726000, Shaanxi,
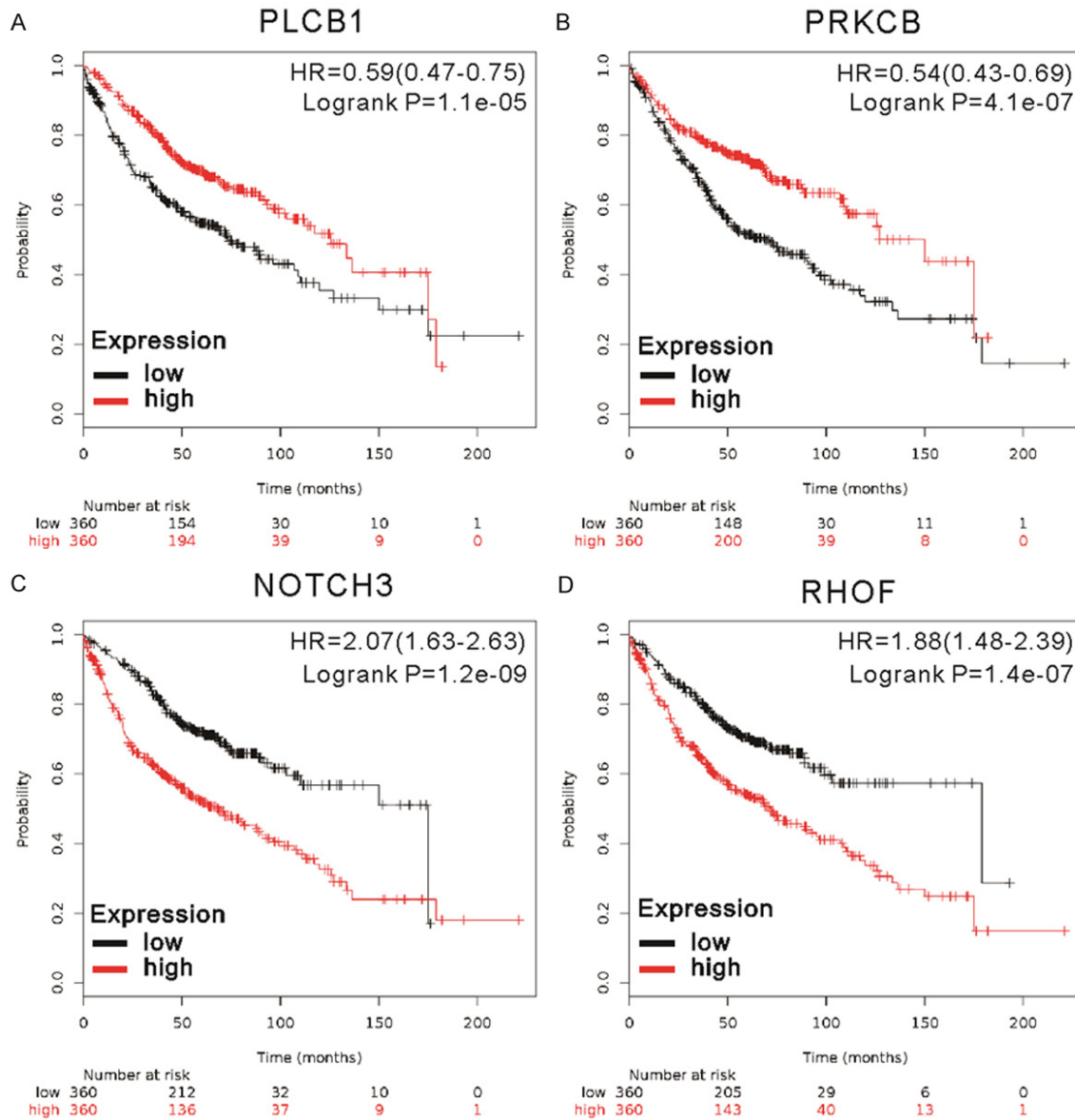
Figure 6. Survival analyses. The prognostic value of each hub genes in NSCLC patients were obtained from KM Plotter and indicated.

China. Tel: +86-0914-2333500; E-mail: yinghongren_vip@163.com

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015; 65: 5-29.
[2] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87-108.
[3] Kris MG, Gaspar LE, Chaft JE, Kennedy EB, Azzoli CG, Ellis PM, Lin SH, Pass HI, Seth R, Shepherd FA, Spigel DR, Strawn JR, Ung YC and Weyant M. Adjuvant systemic therapy and adjuvant radiation therapy for stage i to iiia completely resected non-small-cell lung cancers: American society of clinical oncology/cancer care ontario clinical practice guideline update. J Clin Oncol 2017; 35: 2960-2974.
[4] Xiao L, Lan X, Shi X, Zhao K, Wang D, Wang X, Li F, Huang H and Liu J. Cytoplasmic RAP1 mediates cisplatin resistance of non-small cell lung cancer. Cell Death Dis 2017; 8: e2803.
[5] Hamilton RG. Microarray technology applied to human allergic disease. Microarrays (Basel) 2017; 28; 6.
[6] Akhtar MM, Micolucci L, Islam MS, Olivieri F and Procopio AD. Bioinformatic tools for microRNA dissection. Nucleic Acids Res 2016; 44: 24-44.

[7]   Sun Y, Zheng S, Torossian A, Speirs CK, Schleicher S, Giacalone NJ, Carbone DP, Zhao Z and Lu B. Role of insulin-like growth factor-1 signaling pathway in cisplatin-resistant lung cancer cells. Int J Radiat Oncol Biol Phys 2012; 82: e563-572.

[8]   Nawaz K and Webster RM. The non-small-cell lung cancer drug market. Nat Rev Drug Discov 2016; 15: 229-230.

[9]   Jain A, Jahagirdar D, Nilendu P and Sharma NK. Molecular approaches to potentiate cisplatin responsiveness in carcinoma therapeutics. Expert Rev Anticancer Ther 2017; 17: 815-825.

[10] Ruiz-Ceja KA and Chirino YI. Current FDA-approved treatments for non-small cell lung cancer and potential biomarkers for its detection. Biomed Pharmacother 2017; 90: 24-37.

[11] Schmidmaier R and Baumann P. ANTI-ADHESION evolves to a promising therapeutic concept in oncology. Curr Med Chem 2008; 15: 978-990.

[12] Ciuffreda L, Incani UC, Steelman LS, Abrams SL, Falcone I, Curatolo AD, Chappell WH, Franklin RA, Vari S, Cognetti F, McCubrey JA and Milella M. Signaling intermediates (MAPK and PI3K) as therapeutic targets in NSCLC. Curr Pharm Des 2014; 20: 3944-3957.

[13] Serafin V, Persano L, Moserle L, Esposito G, Ghisi M, Curtarello M, Bonanno L, Masiero M, Ribatti D, Sturzl M, Naschberger E, Croner RS, Jubb AM, Harris AL, Koeppen H, Amadori A and Indraccolo S. Notch3 signalling promotes tumour growth in colorectal cancer. J Pathol 2011; 224: 448-460.

[14] Pasto A, Serafin V, Pilotto G, Lago C, Bellio C, Trusolino L, Bertotti A, Hoey T, Plateroti M, Esposito G, Pinazza M, Agostini M, Nitti D, Amadori A and Indraccolo S. NOTCH3 signaling regulates MUSASHI-1 expression in metastatic colorectal cancer cells. Cancer Res 2014; 74: 2106-2118.

[15] Zhang H, Liu L, Liu C, Pan J, Lu G, Zhou Z, Chen Z and Qian C. Notch3 overexpression enhances progression and chemoresistance of urothelial carcinoma. Oncotarget 2017; 8: 34362-34373.

[16] Gouw LG, Reading NS, Jenson SD, Lim MS and Elenitoba-Johnson KS. Expression of the Rho-family GTPase gene RHOF in lymphocyte subsets and malignant lymphomas. Br J Haematol 2005; 129: 531-533.

[17] Klenke S, Rump K, Buschkamp K, Engler A, Peters J, Siffert W and Frey UH. Characterization of the PLCB1 promoter and regulation by early growth response transcription factor EGR-1. Eur J Pharmacol 2014; 742: 8-14.

[18] Du HF, Ou LP, Yang X, Song XD, Fan YR, Tan B, Luo CL and Wu XH. A new PKCalpha/beta/TBX3/E-cadherin pathway is involved in PLCepsilon-regulated invasion and migration in human bladder cancer cells. Cell Signal 2014; 26: 580-593.

[19] Cirigliano SM, Mauro LV, Grossoni VC, Colombo LL, Diament MJ, Kazanietz MG, Bal de Kier Joffe ED, Puricelli LI and Urtreger AJ. Modulation of pancreatic tumor potential by overexpression of protein kinase C beta1. Pancreas 2013; 42: 1060-1069.