Pre-Clinical Research Report

Journal of INTERNATIONAL MEDICAL RESEARCH

Journal of International Medical Research
48(3) 1–14

DOI: 10.1177/0300060519887637
journals.sagepub.com/home/imr
SAGE

Bioinformatics analysis reveals 6 key biomarkers associated with non-small-cell lung cancer

Bai Dai, Li-qing Ren, Xiao-yu Han and Dong-jun Liu

Abstract

Objective: Non-small-cell lung cancer (NSCLC) accounts for >85% of lung cancers, and its incidence is increasing. We explored expression differences between NSCLC and normal cells and predicted potential target sites for detection and diagnosis of NSCLC.

Methods: Three microarray datasets from the Gene Expression Omnibus database were analyzed using GEO2R. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis were conducted. Then, the String database, Cytoscape, and MCODE plug-in were used to construct a protein–protein interaction (PPI) network and screen hub genes. Overall and disease-free survival of hub genes were analyzed using Kaplan-Meier curves, and the relationship between expression patterns of target genes and tumor grades were analyzed and validated. Gene set enrichment analysis and receiver operating characteristic curves were used to verify enrichment pathways and diagnostic performance of hub genes.

Results: In total, 293 differentially expressed genes were identified and mainly enriched in cell cycle, ECM–receptor interaction, and malaria. In the PPI network, 36 hub genes were identified, of which 6 were found to play significant roles in carcinogenesis of NSCLC: *CDC20*, *ECT2*, *KIF20A*, *MKI67*, *TPX2*, and *TYMS*.

Conclusion: The identified target genes can be used as biomarkers for the detection and diagnosis of NSCLC.

Keywords

Non-small-cell lung cancer, bioinformatic analysis, biomarker, microarray analysis, hub gene, differentially expressed genes

Date received: 20 June 2019; accepted: 20 October 2019

State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, School of Life Sciences, Inner Mongolia University, Hohhot, P. R. China

Corresponding author:
Dong-jun Liu, State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, School of Life Sciences, Inner Mongolia University, Hohhot 010070, P. R. China.
Email: nmliudongjun@126.com

Introduction

Non-small-cell lung cancer (NSCLC) is the most common pathological type of lung cancer and it accounts for more than 85% of all lung cancers.[1] Currently, its morbidity and mortality are increasing from year to year.[2,3] In China, NSCLC is also persistently rising.[4] The occurrence and development of NSCLC are caused by changes in multi-gene expression and various signal transductions.[5,6] As a result, the precise mechanism of NSCLC is difficult to understand. Importantly, early NSCLC-specific symptoms are not obvious and there is no effective diagnostic method for NSCLC in the early stage. Therefore, finding novel biomarkers for diagnosis and prognosis of NSCLC is crucial so that patients can receive appropriate treatment as soon as possible.

In the past years, gene microarray and bioinformatics analysis have been widely used in cancer studies. For instance, Bi et al. and Xu et al.[7,8] identified key genes for diagnosis and treatment of ovarian and bladder cancer by using such methods. Similarly, key biological functions of some genes in the diagnosis and prognosis of NSCLC have been elucidated by means of bioinformatics, such as cyclin-A2 (*CCNA2*), centrosomal protein of 55 kDa (*CEP55*), and neuromedin U (*NMU*).[9,10] The above approach depends on an effective combination of statistics and bioinformatics analysis. However, a separate microarray analysis will increase the false-positive rate of the results.

To minimize the drawbacks of false-positive and false-negative results, we used 3 mRNA microarray datasets in this study to identify target genes affecting NSCLC. We also studied the relationship between the target genes and NSCLC. These identified target genes may be useful for detection and diagnosis of NSCLC.

Materials and methods

Ethical approval

This research did not use animal or human tissue and therefore did not require ethical approval or patient consent.

Microarray data

Gene expression profiles (GSE10072,[11] GSE19804,[12] and GSE43458[13]) were obtained from GEO (http://www.ncbi.nlm.nih.gov/geo), a public functional genomics database containing chip, microarray, and high-throughput gene expression data.[14] The GSE10072 dataset contained 58 NSCLC tissue samples and 49 noncancerous samples; GSE19804 contained 60 NSCLC samples and 60 noncancerous samples; and GSE43458 contained 80 NSCLC samples and 30 noncancerous samples.

Data preprocessing and differential expression analysis

To preprocess the datasets, the differentially expressed genes (DEG) between NSCLC samples and noncancerous samples were screened using GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r). GEO2R is an online network tool in the GEO database that compares DEGs between two groups of samples. LogFC (fold change) >1 and adjusted P-values <0.01 were considered statistically significant.

Enrichment analysis of DEGs

A functional enrichment analysis was performed to examine the enrichment of annotated terms. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses[15] were conducted using DAVID (https://david.abcc.ncifcrf.gov; version 6.7) with a threshold of $P < 0.05$.

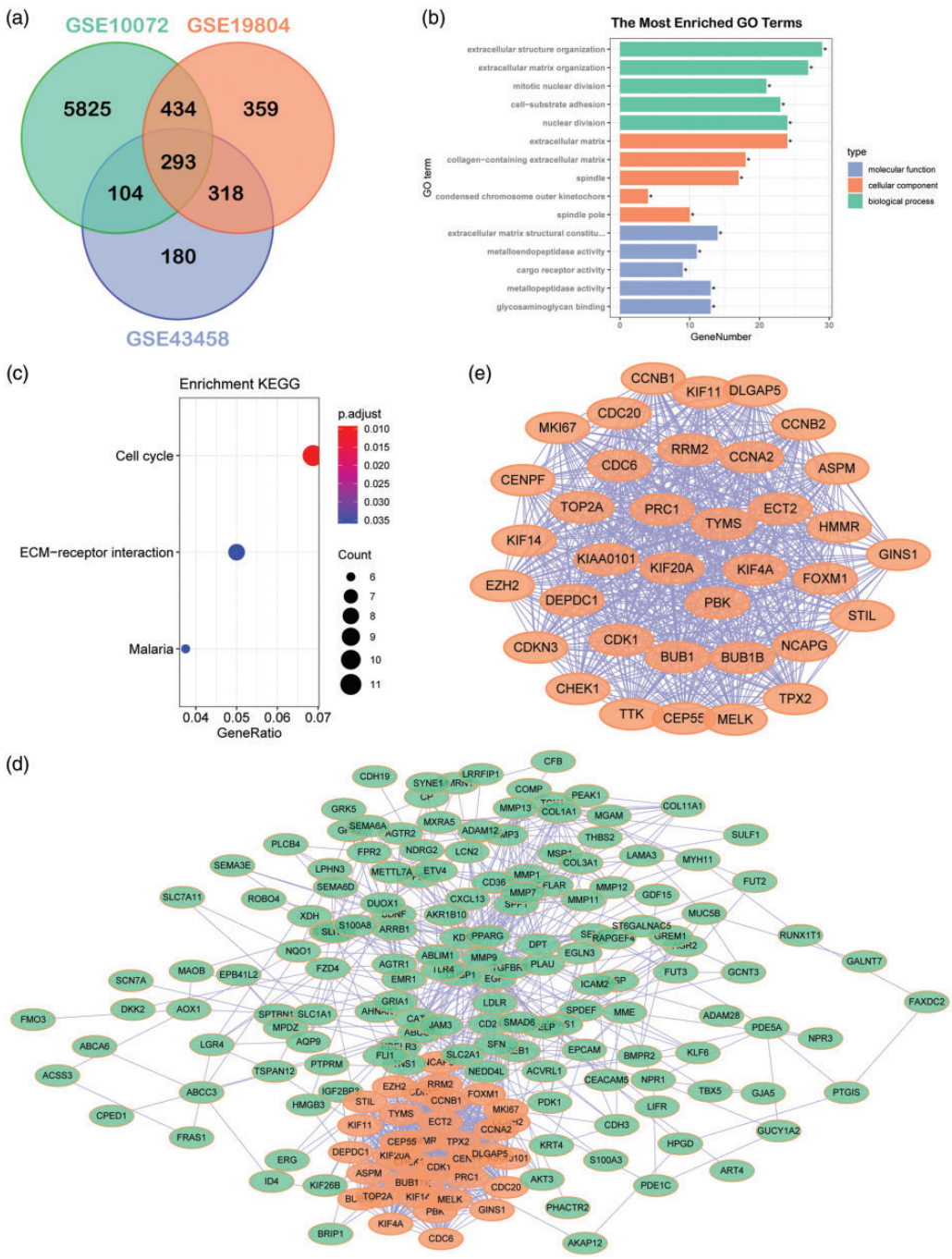

Figure 1. Analysis and screening of DEGs in the PPI network. (a) DEGs with a fold change >2 and *P*-value < 0.01 were selected from among the mRNA expression profiling sets GSE10072, GSE19804, and GSE43458. The 3 datasets showed an overlap of 293 genes. (b) GO and (c) KEGG analysis DEGs. (d) The PPI network of DEGs was constructed using Cytoscape. (e) The most significant module was obtained according MCODE. DEG, differentially expressed gene; PPI, protein–protein interaction; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Protein–protein interaction network construction and cluster analysis

A protein–protein interaction (PPI) network of DEGs was constructed using the String database (http://string-db.org; version 10.0), and an interaction with a combined score >0.4 was considered statistically significant. Subsequently, the results were visualized using Cytoscape,[16] and the most significant module in the PPI network was identified using the MCODE plugin (version 1.5.2). The criteria for selection were as follows: MCODE score >5, degree cut-off = 2, node score cut-off = 0.2, Max depth = 100, and k-score = 2.

Table 1. The list of genes involved in the most significant module.

No.	Gene symbol	Name
01	*CDKN3*	Cyclin dependent kinase inhibitor 3
02	*MELK*	Maternal embryonic leucine zipper kinase
03	*FOXM1*	Forkhead box m1
04	*HMMR*	Hyaluronan mediated motility receptor
05	*KIAA0101*	KIAA0101
06	*NCAPG*	Non-SMC condensin I complex subunit G
07	*GINS1*	GINS complex subunit 1
08	*KIF11*	Kinesin family member 11
09	*CCNB1*	Cyclin B1
10	*CCNA2*	Cyclin A2
11	*CCNB2*	Cyclin B2
12	*BUB1B*	BUB1 mitotic checkpoint serine/threonine kinase B
13	*TPX2*	TPX2, microtubule nucleation factor
14	*KIF4A*	Kinesin family member 4A
15	*BUB1*	BUB1 mitotic checkpoint serine/threonine kinase
16	*RRM2*	Ribonucleotide reductase regulatory subunit M2
17	*DLGAP5*	DLG associated protein 5
18	*TTK*	TTK protein kinase
19	*CHEK1*	Checkpoint kinase 1
20	*ASPM*	Abnormal spindle microtubule assembly
21	*PBK*	PDZ binding kinase
22	*ECT2*	Epithelial cell transforming 2
23	*CDC6*	Cell division cycle 6
24	*CDC20*	Cell division cycle 20
25	*CEP55*	Centrosomal protein 55
26	*KIF14*	Kinesin family member 14
27	*MKI67*	Marker of proliferation Ki-67
28	*CDK1*	Cyclin dependent kinase 1
29	*CENPF*	Centromere protein F
30	*PRC1*	Protein regulator of cytokinesis 1
31	*TYMS*	Thymidylate synthetase
32	*EZH2*	Enhancer of zeste 2 polycomb repressive complex 2 subunit
33	*KIF20A*	Kinesin family member 20A
34	*TOP2A*	Topoisomerase
35	*DEPDC1*	DEP domain containing 1
36	*STIL*	SCL/TAL1 interrupting locus

Target genes scan and analysis

The genes of the module and their co-expressed genes were analyzed using the cBioPortal (http://www.cbioportal.org) online platform.[17,18] The biological process analysis of the genes was performed and visualized using the Biological Networks Gene Oncology tool (BiNGO; version 3.0. 3) plugin of Cytoscape.[19] Hierarchical clustering of hub genes was constructed using the University of California Santa Cruz (UCSC) Cancer Genomics Browser (https://xenabrowser.net/heatmap/).[20] The overall survival and disease-free survival analyses of the genes were assessed using Kaplan-Meier curves in cBioPortal. Furthermore, the survival and receiver operator characteristic (ROC) analyses of hub genes in the TCGA Lung Adenocarcinoma (LUAD) dataset was conducted and visualized using R (www.r-project.org). Gene set enrichment analysis (GSEA) was conducted using GSEA tools (http://www.broad.mit.edu/gsea). The expression profiles of *CDC20*, *ECT2*, *KIF20A*, *MKI67*, *TPX2*, and *TYMS* were analyzed and displayed using the database Oncomine (http://www.oncomine.com). The relationships between expression patterns and tumor grades were also analyzed using Oncomine.[21–23]

Results

Identification of DEGs in NSCLC

After standardization of the microarray results, DEGs (6,656 in GSE10072, 1,404 in GSE19804, and 895 in GSE43458) were

Table 2. Enrichment analysis of DEGs in NSCLC.

Pathway ID	Pathway description	Count	*P*-value	FDR
hsa04110	Cell cycle	10	5.08E-13	3.17E-10
hsa04115	p53 signaling pathway	5	1.34E-05	0.0083229
hsa04914	Progesterone-mediated oocyte maturation	5	3.41E-05	0.0212106
GO:0000279	M phase	20	7.60E-23	1.05E-19
GO:0022403	Cell cycle phase	21	1.29E-22	1.77E-19
GO:0000278	Mitotic cell cycle	20	7.19E-22	9.90E-19
GO:0022402	Cell cycle process	21	6.31E-20	8.69E-17
GO:0000280	Nuclear division	16	7.17E-19	9.87E-16
GO:0007067	Mitosis	16	7.17E-19	9.87E-16
GO:0000087	M phase of mitotic cell cycle	16	9.43E-19	1.30E-15
GO:0007049	Cell cycle	22	1.17E-18	1.61E-15
GO:0048285	Organelle fission	16	1.32E-18	1.81E-15
GO:0051301	Cell division	14	7.52E-14	1.04E-10
GO:0000075	Cell cycle checkpoint	8	1.83E-09	2.51E-06
GO:0051726	Regulation of cell cycle	11	3.77E-09	5.19E-06
GO:0007093	Mitotic cell cycle checkpoint	6	5.67E-08	7.80E-05
GO:0007346	Regulation of mitotic cell cycle	8	6.58E-08	9.06E-05
GO:0006260	DNA replication	7	5.73E-06	0.0078847
GO:0008283	Cell proliferation	9	7.46E-06	0.0102641
GO:0030071	Regulation of mitotic metaphase/anaphase transition	4	1.97E-05	0.0271543
GO:0007017	Microtubule-based process	7	2.92E-05	0.0402219

DEG, differentially expressed gene, NSCLC, non-small-cell lung cancer; FDR, false discovery rate.

identified. The overlap among the 3 datasets contained 293 genes, as shown in the Venn diagram (Figure 1a), consisting of 167 downregulated genes and 126 upregulated genes between NSCLC and noncancerous tissues.

GO and KEGG enrichment analysis of DEGs

The functional and pathway enrichment analysis of DEGs was carried out by using DAVID. The GO analysis results showed

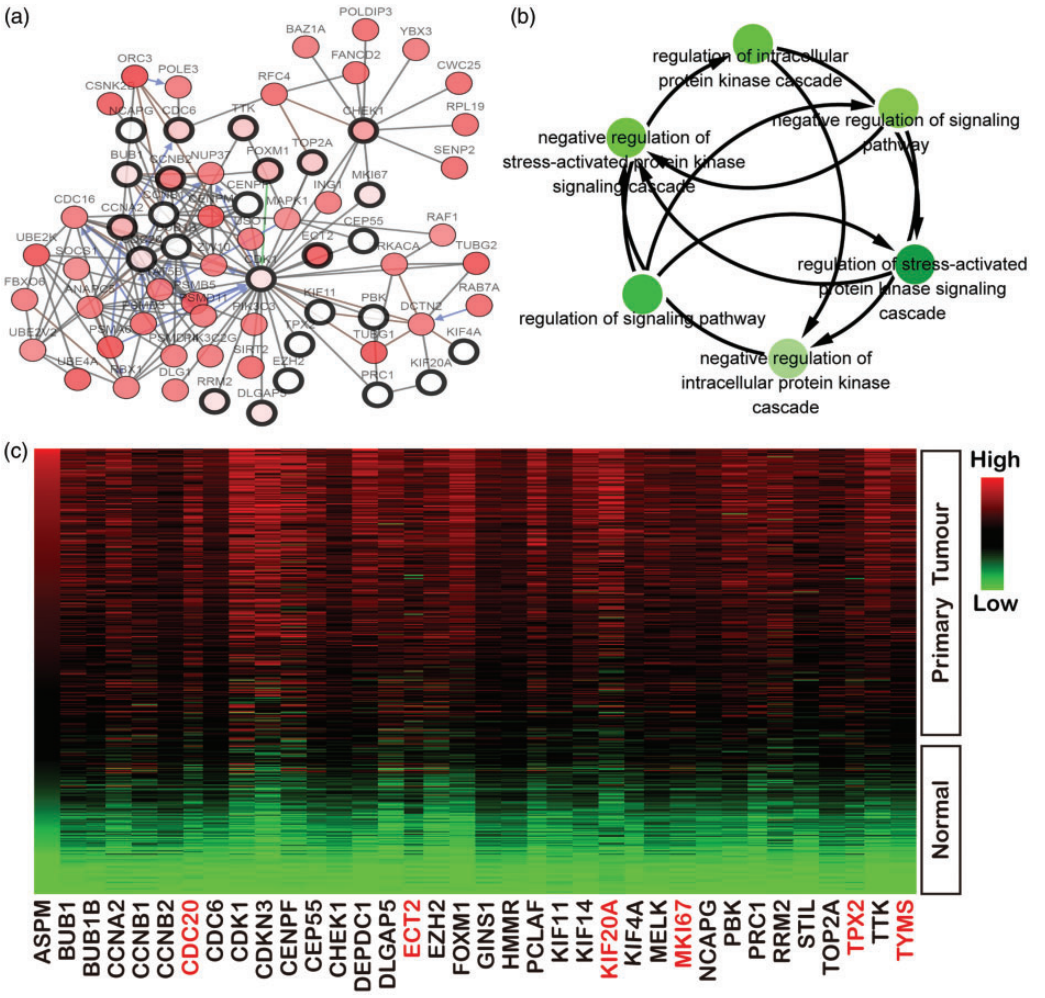

Figure 2. Interaction network and biological process analysis of the module genes. (a) Module genes and their co-expression genes were analyzed using cBioPortal. Nodes outlined in bold black are hub genes; nodes outlined in thin black are co-expression genes. (b) The most significant biological processes of module genes was constructed using BiNGO. (c) Hierarchical clustering of module genes was constructed using UCSC Cancer Genomics Browser (https://xenabrowser.net/heatmap/). Upregulation of genes is marked in red; downregulation of genes is marked in green. Gene symbols shown in red are the six hub genes found to play a significant role in carcinogenesis; gene symbols in black are hub genes identified in the protein–protein interaction network.

that changes in biological processes (BP) of DEGs were significantly enriched in extracellular matrix (ECM) structural constituent, metalloendopeptidase activity, cargo receptor activity, metallopeptidase activity, and glycosaminoglycan binding. Changes in molecular function (MF) were mainly enriched in extracellular structure organization, ECM organization, mitotic nuclear division, cell substrate adhesion, and nuclear division. Changes in cell component (CC) of DEGs were mainly enriched in the ECM, collagen containing ECM, spindle, condensed chromosome outer kinetochore, and spindle pole (Figure 1b). KEGG pathway analysis revealed that the DEGs were mainly enriched in cell cycle, ECM–receptor interaction, and malaria (Figure 1c).

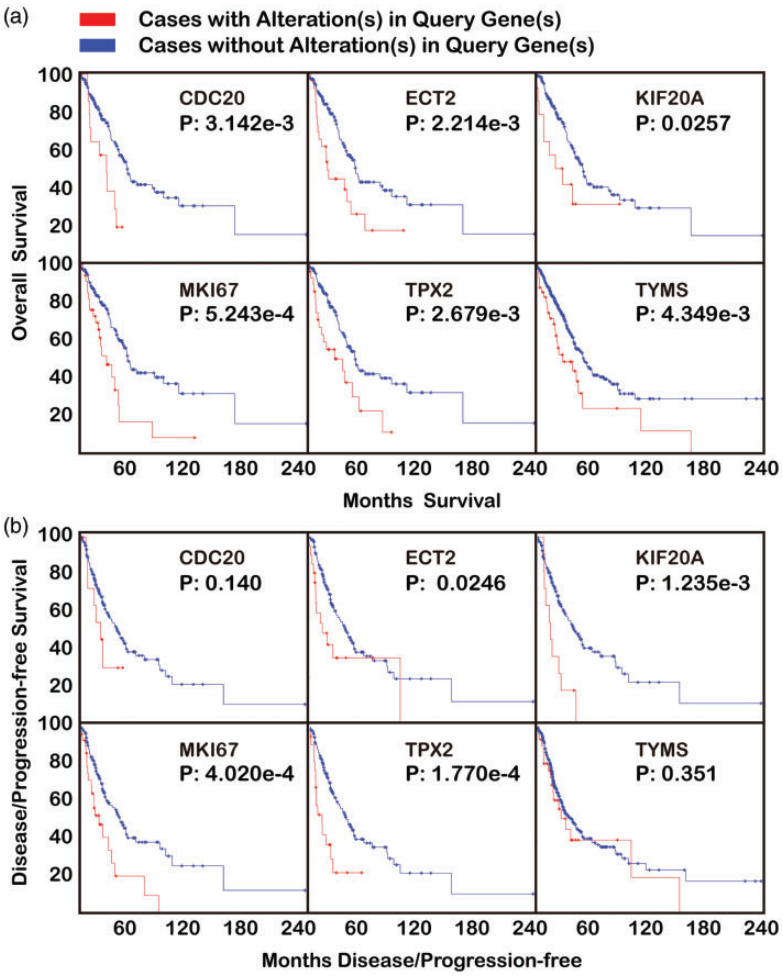

Figure 3. Survival analysis of potential NSCLC biomarkers. (a) Overall survival, and (b) disease-free survival analyses of module genes were analyzed using cBioPortal online platform. $P < 0.05$ was considered statistically significant. NSCLC, non-small-cell lung cancer.

Construction and module analysis of PPI Network

The PPI network of DEGs is shown in Figure 1d, and the most significant module is shown in Figure 1e. The abbreviations and full names for genes involved in this module are shown in Table 1. The functional analyses of these genes showed that 36 genes in this module were mainly enriched in cell division, mitotic nuclear division, and cell cycle (Table 2). A network of these genes and their co-expressed genes is shown in Figure 2a. The most significant

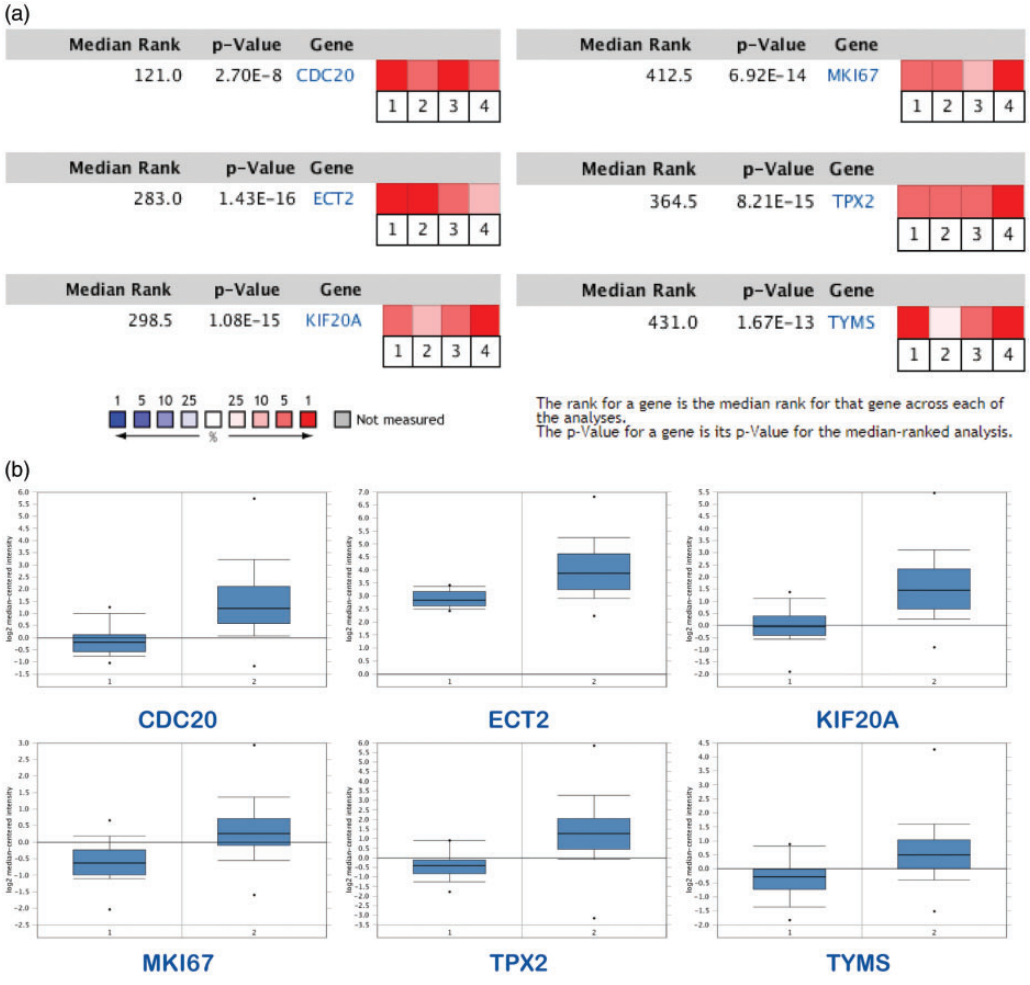

Figure 4. Oncomine analysis of NSCLC samples and noncancerous samples of potential NSCLC biomarkers. (a) Heat maps of potential NSCLC biomarker expression in clinical lung carcinoma samples versus normal tissues. 1 = Lung Adenocarcinoma vs. Normal Landi Lung, PLoS ONE, 2008. 2 = Lung Adenocarcinoma vs. Normal Okayama Lung, Cancer Res, 2012. 3 = Lung Adenocarcinoma vs. Normal Selamat Lung, Genome Res, 2012. 4 = Lung Adenocarcinoma vs. Normal Su Lung, BMC Genomics, 2007. (b) mRNA expression in NSCLC compared with normal lung tissues. Lower and upper circles indicate the minimum and maximum values, whiskers indicate the 10th and 90th percentiles, and the box indicates the 25th and 75th percentiles, respectively; the line indicates the median. NSCLC, non-small-cell lung cancer.

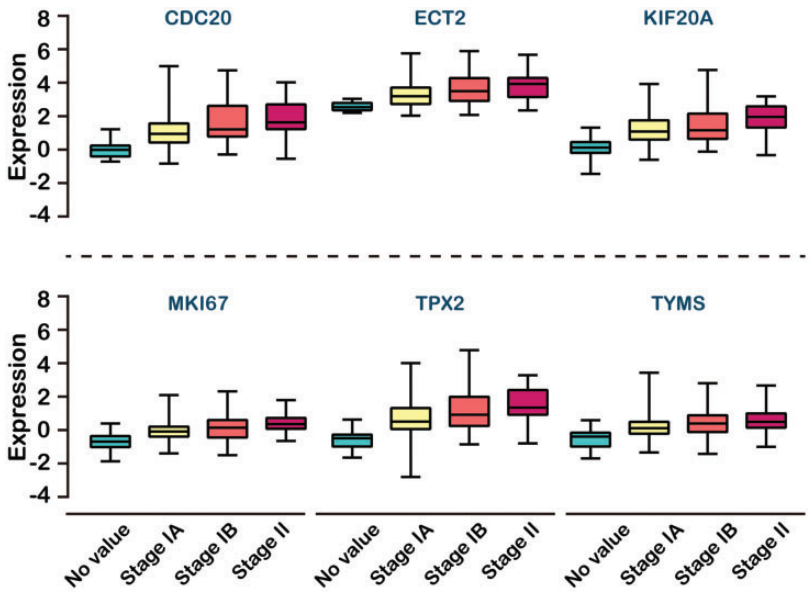

Figure 5. Association between the expression of potential NSCLC biomarkers and tumor stage. NSCLC, non-small-cell lung cancer.

biological processes of these genes is shown in Figure 2b. Subsequently, hierarchical clustering showed that the hub genes could differentiate the NSCLC samples from the noncancerous samples (Figure 2c).

Analysis of potential biomarkers for NSCLC

NSCLC patients with alterations in *CDC20*, *ECT2*, *MKI67*, *TPX2*, and *TYMS* showed worse overall survival (Figure 3a), and NSCLC patients with alterations in *KIF20A*, *MKI67*, and *TPX2* showed worse disease-free survival (Figure 3b). Therefore, these genes can be identified as potential NSCLC biomarkers. The Oncomine analysis of cancerous versus normal tissue showed that these genes were significantly overexpressed in NSCLC in the different datasets (Figure 4a and b). Meanwhile, higher mRNA expression of these genes was associated with tumor stage in the Oncomine lung datasets (Figure 5). To clarify the accuracy of this result, we validated these genes by using the TCGA database (Figure 6), and based on the TCGA database, we validated the GSEA. The gene sets with the highest enrichment scores were all closely associated with cell cycle (Figure 7). In addition, ROC curves showed that all these genes could serve as biomarkers to distinguish tumors from normal lung tissue sensitively and accurately. All these genes appeared to be promising candidates for therapeutic targets (Figure 8).

Discussion

Biomarkers for diagnosing or treating cancer are often obtained by identifying the most important DEGs in microarray or high-throughput case-control studies.[7] As with any other cancer, the development, progression, and metastasis of lung cancer is a very complex process, involving multiple gene and cellular pathway aberrations.[24] The DEGs between NSCLC and normal

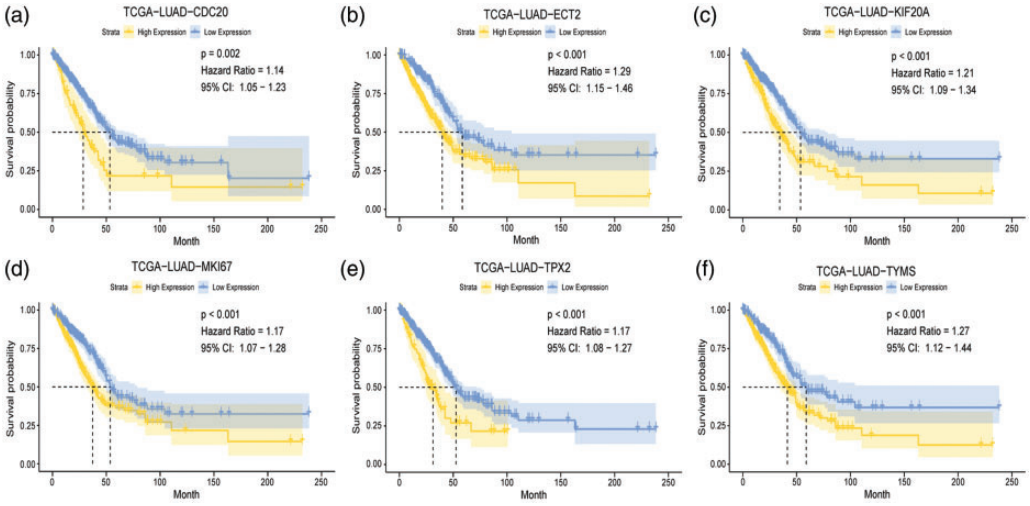

Figure 6. Survival analysis of potential NSCLC biomarkers using TCGA database. Analyses of CDC20 (a), ECT2 (b), KIF20A (c), MKI67 (d), TPX2 (e), and TYMS (f) were carried out. NSCLC, non-small-cell lung cancer.

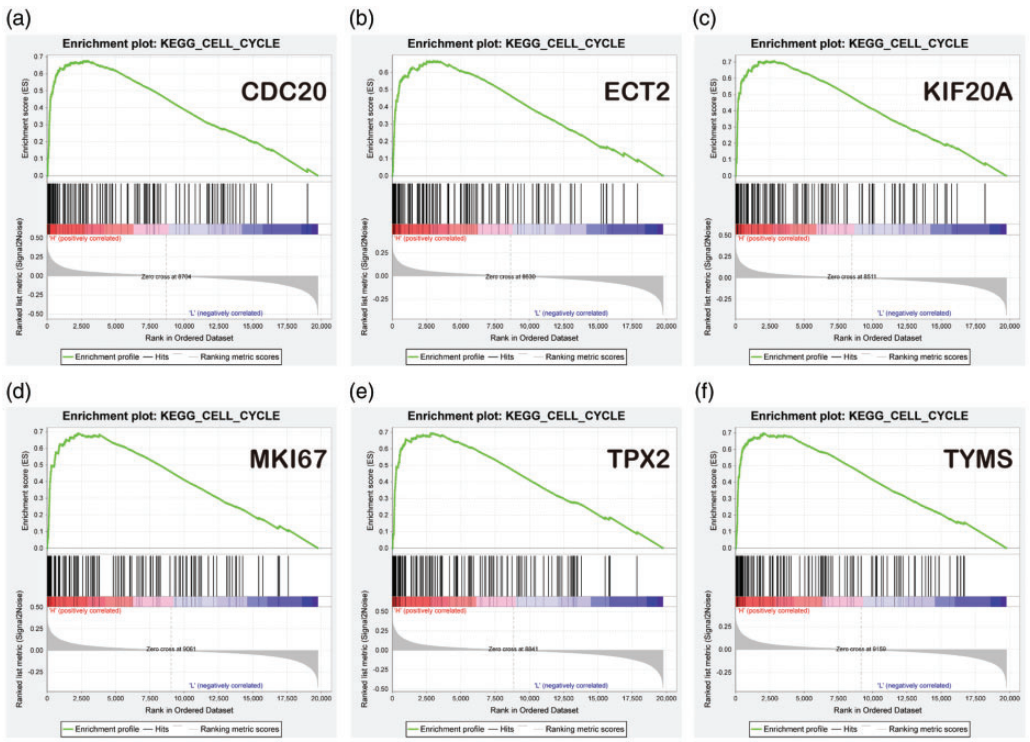

Figure 7. Gene set enrichment analysis of potential NSCLC biomarkers using the TCGA database. Analyses of CDC20 (a), ECT2 (b), KIF20A (c), MKI67 (d), TPX2 (e) and TYMS (f) were carried out. NSCLC, non-small-cell lung cancer.

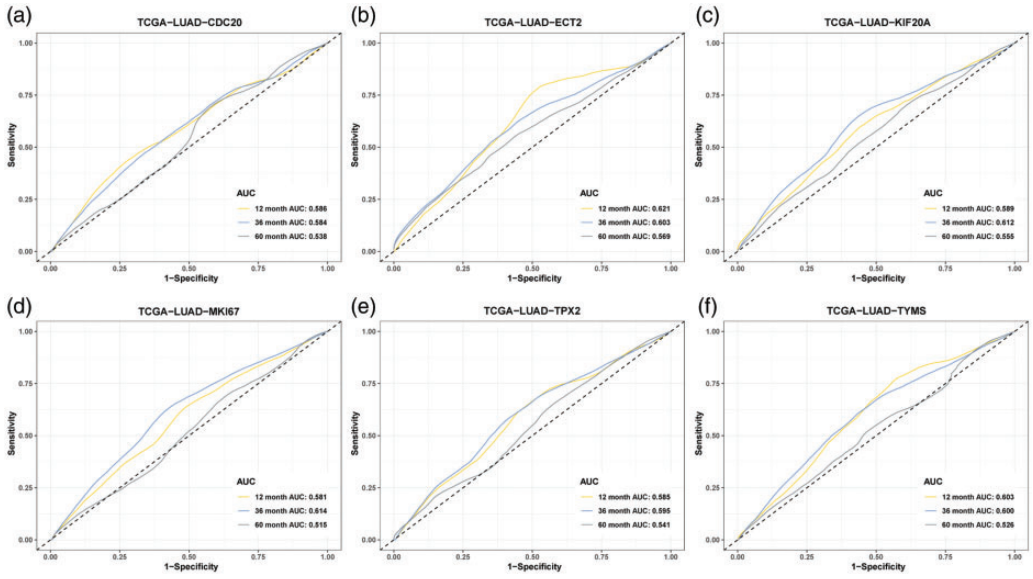

Figure 8. Receiver operating characteristic curve analysis of potential NSCLC biomarkers using the TCGA database. Analyses of CDC20 (a), ECT2 (b), KIF20A (c), MKI67 (d), TPX2 (e), and TYMS (f) were carried out. NSCLC, non-small-cell lung cancer.

tissue may be the core functional genes that promote the occurrence and development of NSCLC.[25,26] To improve the diagnosis and treatment of NSCLC, it is important to identify these DEGs and understand their role in the molecular mechanisms of NSCLC.

In the present study, 293 DEGs were identified between NSCLC and noncancerous samples through analysis of three datasets. Among these DEGs, we selected 6 that are closely related to the occurrence and development of NSCLC: *CDC20*, *ECT2*, *KIF20A*, *MKI67*, *TPX2*, and *TYMS*. When the overall survival and disease-free survival analyses of target genes were performed, we found that poor prognosis of NSCLC patients was associated with high expression of target genes. Kato et al.[27] reported that *CDC20* was overexpressed in NSCLC, and that overexpression predicts poor prognosis. Bai et al.[28] showed that the overexpression of *ECT2* could promote the occurrence and development of NSCLC, suggesting that *ECT2* could be used as a diagnostic marker. Ni et al.,[29] using bioinformatics analysis, showed that *KIF20A* was correlated with the pathogenesis and prognosis of NSCLC. Schneider et al.[30] demonstrated that overexpression of *TPX2* mRNA in tumor cells is associated with the prognosis of NSCLC patients. Sun et al.[31] showed that mRNA expression of *TYMS* may have prognostic value for patients with NSCLC treated with platinum-based chemotherapy. These previous studies are consistent with our results and demonstrate the effectiveness of bioinformatics in screening to identify target genes. However, we found no reports associating *MKI67* with NSCLC. Therefore, the function of *MKI67* to NSCLC needs further experimental confirmation.

In our study, we identified 36 hub genes. Hub genes are involved in many biological processes and induce many signal transductions. Therefore, analyzing the biological functions and signaling pathways related

to hub genes can effectively reveal the occurrence and development of NSCLC. GO enrichment analysis revealed that hub genes were mainly enriched in extracellular structure organization, ECM organization, mitotic nuclear division, cell substrate adhesion, and nuclear division, whereas changes in KEGG were mainly enriched in cell cycle, ECM-receptor interaction, and malaria. Previous studies have reported that dysregulation of the cell cycle plays an important role in the carcinogenesis or progression of tumors.[32,33] CDC20 can act as a regulatory protein that interacts with other proteins to participate in the cell cycle of tumors.[34] CDC20 has also been shown to be involved in tumor formation by regulating the ECM-receptor interaction pathway.[35] These studies are consistent with our research on CDC20 and confirm our results. However, a large number of studies in NSCLC still need to be further explored.

In conclusion, our research objective was to find new biomarkers related to the diagnosis and prognosis of NSCLC. A total of 293 DEGs and 36 hub genes were identified, and 6 target genes closely related to NSCLC were identified by screening. These bioinformatics analyses provide a new perspective to further understand the occurrence and development of NSCLC and have a positive effect on the treatment of NSCLC. However, the results still need to be rigorously evaluated before clinical treatment can be performed.

Acknowledgements

Special thanks go to Xiu-juan Chen (Reproductive Medicine Center, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, P. R. China), Ri-na Sha (Pathology Department, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, P. R. China), Jian-long Yuan (Clinical Laboratory, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, P. R. China), and Jie Zhao (Reproductive Medicine Center, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, P. R. China) for suggestions with the paper.

Author contributions

Dong-jun Liu designed and supervised the research. Bai Dai performed statistical analyses and wrote the manuscript. Li-qing Ren and Xiao-yu Han did the practical work and revised the manuscript.

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

This work was supported by the Science and Technology Innovation Guided Project in Inner Mongolia Autonomous Region (KCBJ2018003).

ORCID iD

Dong-jun Liu https://orcid.org/0000-0002-5629-5931

References

1. Jemal A, Thun M, Yu XQ, et al. Changes in smoking prevalence among U.S. adults by state and region: estimates from the Tobacco Use Supplement to the Current Population Survey, 1992-2007. *BMC Public Health* 2011; 11: 512.
2. Field JK, Hansell DM, Duffy SW, et al. CT screening for lung cancer: countdown to implementation. *Lancet Oncol* 2013; 14: e591–e600.
3. Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin* 2011; 61: 69–90.
4. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016; 66: 115–132.
5. Kalari KR, Rossell D, Necela BM, et al. Deep sequence analysis of non-small cell lung cancer: integrated analysis of gene expression, alternative splicing, and single nucleotide variations in lung adenocarcinomas with and without oncogenic KRAS mutations. *Front Oncol* 2012; 2: 12.

6. Markou A, Sourvinou I, Vorkas PA, et al. Clinical evaluation of microRNA expression profiling in non small cell lung cancer. *Lung Cancer* 2013; 81: 388–396.
7. Bi D, Ning H, Liu S, et al. Gene expression patterns combined with network analysis identify hub genes associated with bladder cancer. *Comput Biol Chem* 2015; 56: 71–83.
8. Xu Z, Zhou Y, Cao Y, et al. Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis. *Med Oncol* 2016; 33: 130.
9. Xiao Y, Feng M, Ran H, et al. Identification of key differentially expressed genes associated with nonsmall cell lung cancer by bioinformatics analyses. *Mol Med Rep* 2018; 17: 6379–6386.
10. You S and Gao L. Identification of NMU as a potential gene conferring alectinib resistance in non-small cell lung cancer based on bioinformatics analyses. *Gene* 2018; 678: 137–142.
11. Landi MT, Dracheva T, Rotunno M, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* 2008; 3: e1651.
12. Lu TP, Tsai MH, Lee JM, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev* 2010; 19: 2590–2597.
13. Kabbout M, Garcia MM, Fujimoto J, et al. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. *Clin Cancer Res* 2013; 19: 3383–3395.
14. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 2009; 37: D885–D890.
15. Huang DW, Sherman BT, Tan Q, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 2007; 8: R183.
16. Lopes CT, Franz M, Kazi F, et al. Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 2010; 26: 2347–2348.
17. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; 2: 401–404.
18. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; 6: pl1.
19. Maere S, Heymans K and Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005; 21: 3448–3449.
20. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002; 12: 996–1006.
21. Chen X, Cheung ST, So S, et al. Gene expression patterns in human liver cancers. *Mol Biol Cell* 2002; 13: 1929–1939.
22. Roessler S, Jia HL, Budhu A, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res* 2010; 70: 10202–10212.
23. Wurmbach E, Chen YB, Khitrov G, et al. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* 2007; 45: 938–947.
24. Procaccio L, Bergamo F, Manai C, et al. An overview on clinical, pathological and molecular features of lung metastases from colorectal cancer. *Expert Rev Respir Med* 2019; 13: 635–644.
25. Bosse Y and Amos CI. A decade of GWAS results in lung cancer. *Cancer Epidemiol Biomarkers Prev* 2018; 27: 363–379.
26. Cheung CHY and Juan HF. Quantitative proteomics in lung cancer. *J Biomed Sci* 2017; 24: 37.
27. Kato T, Daigo Y, Aragaki M, et al. Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients. *J Surg Oncol* 2012; 106: 423–430.
28. Bai X, Yi M, Xia X, et al. Progression and prognostic value of ECT2 in non-small-cell lung cancer and its correlation with PCNA. *Cancer Manag Res* 2018; 10: 4039–4050.
29. Ni M, Liu X, Wu J, et al. Identification of candidate biomarkers correlated with the

pathogenesis and prognosis of non-small cell lung cancer via integrated bioinformatics analysis. *Front Genet* 2018; 9: 469.
30. Schneider MA, Christopoulos P, Muley T, et al. AURKA, DLGAP5, TPX2, KIF11 and CKAP5: five specific mitosis-associated genes correlate with poor prognosis for non-small cell lung cancer patients. *Int J Oncol* 2017; 50: 365–372.
31. Sun S, Shi W, Wu Z, et al. Prognostic significance of the mRNA expression of ERCC1, RRM1, TUBB3 and TYMS genes in patients with non-small cell lung cancer. *Exp Ther Med* 2015; 10: 937–941.
32. Moyer TC and Holland AJ. PLK4 promotes centriole duplication by phosphorylating STIL to link the procentriole cartwheel to the microtubule wall. *Elife* 2019; 8: pii: e46054.
33. Shah H, Rawat K, Ashar H, et al. Dual role for fungal-specific outer kinetochore proteins during cell cycle and development in *Magnaporthe oryzae*. *J Cell Sci* 2019; 132: pii: jcs224147.
34. Cheng S, Castillo V and Sliva D. CDC20 associated with cancer metastasis and novel mushroom-derived CDC20 inhibitors with antimetastatic activity. *Int J Oncol* 2019; 54: 2250–2256.
35. Lv M and Wang L. Comprehensive analysis of genes, pathways, and TFs in nonsmoking Taiwan females with lung cancer. *Exp Lung Res* 2015; 41: 74–83.