Research Article

Identification of CDKN3 and UBE2C mRNAs as Prognostic Biomarkers in Early-Stage Lung Adenocarcinoma Using Bioinformatics Strategy

Qiang Chen [1, 2, 3, *], Lutong Xu [1], Jing Hu [1, 2, 3], Tonglian Wang [1], Kang Zhang [1], Hongbo Zhao [4], Yuanyue Li [1] and Tao Shou [2, 3]

[1]Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, China
[2]Medical Oncology, The First People's Hospital of Yunnan Province, Kunming, China
[3]Medical Oncology, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, China
[4]Institute of Molecular and Clinical Medicine, Kunming Medical University, Kunming, China

[*]Corresponding author: Faculty of Life Science and Technology, Kunming University of Science and Technology, 650500, Kunming, China. Tel: +86-87165939528, Email: chq@sjtu.edu.cn

Abstract

Background: Lung adenocarcinoma (LUAD) is the most common histological subtype of non-small cell lung cancer with very poor 5-year overall survival (OS) rate. It is histopathologically difficult to predict clinical outcome in early-stage LUAD. Identifying reliable prognostic biomarker is absolutely critical to benefit from early additional treatment for early-stage LUAD patients.

Objectives: The purpose of the current study was to identify critical genes as prognostic biomarkers in early-stage LUAD using gene expression profiles based on the microarray.

Methods: In this bioinformatics-based cross-study, gene expression profiles from early-stage LUAD, including GSE10072 and GSE19804 genes were integrated using bioinformatics methods, including differentially expressed gene analysis (DEGA), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, and protein-protein interaction (PPI) network construction. Subsequently, the survival analysis of key genes was performed using The Cancer Genome Atlas (TCGA) database and was validated using online Gene Expression Profiling Interactive Analysis (GEPIA) database.

Results: A total of 89 up-regulated and 214 down-regulated genes were identified in early-stage LUAD, and the functional changes of 303 differentially expressed genes (DEGs) were mainly related to cell cycle. A PPI network was established by online STRING database with 207 nodes and 775 edges. Centrality analysis showed that *CDKN3* and *UBE2C* genes were identified as key genes implicated in early-stage LUAD. Survival analysis revealed that low mRNA expressions of *CDKN3* and *UBE2C* were significantly associated with longer OS of early-stage LUAD patients.

Conclusions: This cross-study found key dysregulated genes involved in early-stage LUAD, which might provide insights into the pathogenesis of early-stage LUAD, and identified *UBE2C* and *CDKN3* might serve as potential diagnostic and prognostic biomarkers and therapeutic targets for early-stage LUAD.

Keywords: Adenocarcinoma, Bioinformatics, Biomarker, Early, Gene, Lung, Outcome, Prognostic, Stage, Survival

1. Background

Lung cancer is the leading cause of cancer-related death worldwide and results in more than 1.3 million deaths annually (1). Non-small cell lung cancer (NSCLC) is the most important pathological type and accounts for about 85% of all lung cancer cases (2). Despite recent advances in multi-modality diagnosis and therapy, the majority of NSCLC cases are diagnosed at an advanced stage (III or IV stage)(3), and the overall 5-year and 10-year survival rates are only unoptimistically 17% and pessimistically 8-10% (4). Given the difficulties in the treatment of advanced NSCLC, the most promising way to improve outcomes may be an effective diagnosis and treatment of early-stage NSCLC patients (5). Indeed, the early-stage efficient diagnosis of NSCLC contribute to offer a favorable prognosis and the overall 5-year survival rate will increase to 70-90% (6). Currently, disease stage and histological grade are the basis to evaluate NSCLC diagnosis and prognosis. However, clinical and pathological symptoms usually limit predictive value in detecting early NSCLC, and clinical outcomes are highly variable due to the heterogeneity of NSCLC. Therefore, it is vital to identify potential diagnostic and prognostic biomarkers and/or therapeutic targets for combating NSCLC.

Lung adenocarcinoma (LUAD) is the most common his-

tological subtype of NSCLC (7), resulting in more than 40% lung cancer death per year and the morbidity and mortality are increasing year by year (8). Although researches have shown that smoking-tobacco accelerated LUAD development, LUAD indicates the lowest association with smoking-tobacco among all histological types and gene aberrations often play key roles in triggering LUAD (9). Despite many gene aberrations during LUAD development, LUAD is often triggered by an aberration of a driver gene (7). Gene expression analysis is the most common tool to identify differentially expressed genes (DEGs) between tumor and normal tissues. Using gene expression profiles, hundreds of LUAD-related DEGs were detected, including some key driver genes such as epidermal growth factor receptor (*EGFR*) and anaplastic lymphoma kinase (*ALK*) (10-13), and some gene expression signatures were found as prognostic biomarkers (5). However, due to the heterogeneity of LUAD pathogenesis, those prognosticators have been not widely accepted, and reliable, consistent prognosticators based on gene expression need further elucidation (5).

2. Objectives

The increasing available LUAD data makes it possible to search consistent gene expression signatures. In this study, two early-stage LUAD-related gene expression profiles, including GSE10072 and GSE19804 from NCBI GEO database were integrated to detect DEGs involved in early-stage LUAD. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways related to the identified DEGs were investigated. A protein and protein interaction (PPI) network of DEGs encoding proteins was constructed to elucidate the interactive relationships among DEGs and centrality analysis was used to identify key genes. Survival analysis of key genes was performed to detect the associations of key genes and overall survival (OS) of early-stage LUAD patients. Gene Expression Profiling Interactive Analysis (GEPIA) database was used to validate key genes related to OS of LUAD patients.

3. Methods

3.1. Gene Expression Data Collection

In this bioinformatics-based cross-study, two gene expression profiles associated with LUAD, including GSE10072 and GSE19804 were retrieved from the NCBI Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). GSE10072 data was from American LUAD patients and consisted of 107 samples that contained 58 LUAD samples and 49 normal samples (11). GSE10072 data was produced using Affymetrix Human

Genome U133A Array platform. GSE19804 data was from Taiwanese LUAD patients and contained 60 LUAD samples and 60 normal samples (12). GSE19804 data was generated using the Affymetrix Human Genome U133 Plus 2.0 Array platform. Because two expression profiles contained the data of all stages, the middle- and late-stage LUAD data were removed, and the early-stage LUAD data was kept. Finally, 83 samples (43 early-stage LUAD samples and 40 normal samples) for GSE10072 and 94 samples (47 early-stage LUAD samples and 47 normal samples) for GSE19804 were used.

In this study, publicly available gene expression profiles from the NCBI GEO database were collected with patients' consent approved by the relevant institutional review board. GSE10072 data was approved by the Institutional Review Board of the relevant participating hospital and by the National Cancer Institute (Bethesda, MD) (11). GSE19804 data was approved by the Institutional Review Board of Taiwan University Hospital and by the Institutional Review Board of Taichung Veterans General Hospital (12). The present study met the requirements of data usage and publishing from the NCBI GEO database.

3.2. Data Preprocessing and DEGs Screening

All raw data were standardized by a normalized microarray preprocessing procedure using affy package (version 1.60.0) in Bioconductor project (version 3.7.0, http://www.bioconductor.org/) to eliminate the expression change caused by the experimental technique (14). Differentially expressed gene analysis (DEGA) was performed using the limma package (version 3.36.1) based empirical Bayes method in the Bioconductor project to screen DEGs (15). DEGs between LUAD and normal samples were identified according to |log2 (fold change) | (|logFC|) > 1 and false discovery rate (FDR) < 0.05 cutoff criteria. Commonly dysregulated genes between GSE10072 and GSE19804 were used for subsequent analyses.

3.3. KEGG Pathway Enrichment Analysis

KEGG pathway enrichment analysis of commonly dysregulated genes was performed using the clusterProfiler package (version 3.10.0) in Bioconductor project (16) and a KEGG pathway with an adjusted P < 0.05 was considered to be statistically significant.

3.4. PPI Network Construction and Module Analysis

The interactive relationships among common DEGs encoding proteins were analyzed by constructing a PPI network. The interactive information among DEGs was obtained by online STRING database (version 10.5, https://string-db.org/) (17). The gene pairs with combined

scores > 0.4 were used for PPI network construction. Cytoscape software (version 3.6.1, http://www.cytoscape.org/) was used to construct and visualize the PPI network (18).

To detect highly interconnected clusters (PPI subnetwork) within the PPI network, topological properties were analyzed using Molecular COmplex DEtection (MCODE) algorithm, and a plugin MCODE (version 1.4.1) in Cytoscape was used to perform MCODE analysis (19). The threshold parameters were set for maximum depth = 100, node score = 0.2, and K-core = 2.

3.5. Key Gene Identification and Validation

Centrality analysis is a principal method for identifying key DEGs encoding proteins in PPI network and CytoNCA app (version 2.1.6) in Cytoscape was used to perform centrality analysis (20). Three centrality methods in centrality analyses, including Subgraph centrality, Degree centrality, and Closeness centrality were used to identify genes with higher PPI scores (20). Key genes were identified as the intersecting genes of the genes obtained by three centrality methods.

The early-stage LUAD data from The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/) was used to evaluate the associations between key genes and OS in early-stage LUAD patients were estimated using Kaplan-Meier (KM) estimate and Log-rank (LR) test in survival (version 2.43-3) package in R project. A gene with statistical P < 0.05 was considered the significant association between the gene and OS. GEPIA database (http://gepia.cancer-pku.cn) was an interactive web server for analyzing gene expression data of tumors and normal tissues from TCGA and genotype-tissue expression database (21) and was used to validate key genes related to OS of LUAD patients. Survival curve and boxplot were used to visualize the relationships. Pearson correlation analysis was used to detect the correlation of expression pattern between key genes related to OS. The statistical P < 0.05 was considered a significant correlation between two genes in expression pattern.

3.6. Statistical Analysis

The comparison of expression of the key gene in LUAD tissues and normal tissues was performed using the mean ± standard deviation. Statistical differences were estimated using *t*-test based on R language. P < 0.05 was considered a statistically significant difference.

4. Results

4.1. DEGs Identification

We used |logFC| > 1 and FDR < 0.05 as detecting criteria for screening DEGs between early-stage LUAD samples

and normal samples from GSE10072 and GSE19804. Subsequently, 503 DEGs, including 175 up-regulated DEGs and 328 down-regulated DEGs were extracted in GSE10072. Also, 745 DEGs, including 233 up-regulated DEGs and 512 down-regulated DEGs were screened in GSE19804. An overlapping analysis showed that 89 up-regulated and 214 down-regulated genes were identified.

4.2. KEGG Pathway Evaluation

To better understand the roles of identified common DEGs in early-stage LUAD, KEGG pathway enrichment analysis of common DEGs was performed. According to an adjusted P < 0.05, 6 dysregulated pathways were found to be significantly enriched (Table 1). Among them, 3 pathways were enriched by 89 up-regulated DEGs and were protein digestion and absorption (hsa04974), ECM-receptor interaction (hsa04512), and cell cycle (hsa04110). The other 3 pathways were enriched by 214 down-regulated DEGs, and were complement and coagulation cascades (hsa04610), renin-angiotensin system (hsa04614), and malaria (hsa05144).

4.3. PPI Network Construction and Module Identification

The interactive relationships among common DEGs encoding proteins were elucidated using the PPI network and the interactive information among DEGs was obtained from online STRING database. At a combined score > 0.4, a total of 207 DEGs (67 up-regulated and 140 down-regulated) among 303 common DEGs was filtered into the PPI network with 207 nodes and 775 edges (Figure 1A). Highly correlated module analysis showed that 12 PPI modules were found in the PPI network and the most significant PPI module was comprised of 22 nodes with 230 edges (Figure 1B). Node degree analysis showed that 20 nodes among 22 nodes interacted with each other and had closer relationships. Significantly enriched 5 genes within the cell cycle pathway, including *BUB1B*, *CCNB1*, *CDC20*, *PTTG1*, and *TTK* were observed to exist in the most significant PPI module.

4.4. Key Gene Identification

Centrality analyses were used to identify key genes involved in early-stage LUAD. Based on the three centrality methods, including Subgraph centrality, Degree centrality, and Closeness centrality, Top 20 genes obtained by each method were selected as key candidate genes (Table 2). An overlapping analysis showed that 3 genes, including cyclin-dependent kinase inhibitor 3 (*CDKN3*, logFC = 1.24, and P = 1.39e-12 in GSE10072, logFC = 1.33 and P = 7.32e-09 in GSE19804), ubiquitin-conjugated enzyme E2 (*UBE2C*, logFC = 1.25, and P = 1.62e-11 in GSE10072, logFC = 1.34 and P = 6.69e-12 in GSE19804), and enhancer of zeste homolog 2 (*EZH2*,

Table 1. Pathway Enrichment Analysis of DEGs Function in Early-Stage LUAD

Pathway ID	Description	Adjusted P Value	Count	Gene Symbol
Up-regulated				
hsa04974	Protein digestion and absorption	8.20e-5	7	*COL10A1, COL11A1, COL1A1, COL1A2, COL3A1, COL5A1, COL5A2*
hsa04512	ECM-receptor interaction	5.23e-3	5	*COL1A1, COL1A2, COMP, SPP1, THBS2*
hsa04110	The cell cycle	2.37e-2	5	*BUB1B, CCNB1, CDC20, PTTG1, TTK*
Down-regulated				
hsa04610	Complement and coagulation cascades	3.36e-6	11	*C4BPA, C5AR1, C7, CFD, CPB2, F8, PROS1, SERPING1, THBD, VSIG4, VWF*
hsa04614	Renin-angiotensin system	3.01e-2	4	*AGTR1, AGTR2, CPA3, MME*
hsa05144	Malaria	4.58e-2	5	*ACKR1, CD36, HBB, IL6, PECAM1*

Abbreviation: LUAD, lung adenocarcinoma.

logFC = 1.09, and P = 5.72e-09 in GSE10072, logFC = 1.11 and P = 1.55e-07 in GSE19804) were intersecting genes of top 20 genes obtained by three methods.

Compared with the 22 genes in the most significant PPI module, those 3 intersecting genes were included in the most significant PPI module (Figure 1C), and had closer relationships with 5 genes enriched within cell cycle pathway (Figure 1B). However, the combined scores and co-expression scores of *EZH2* and those 5 genes were the lowest. The PPI network based on each gene confirmed the result. At a maximum number of interactors ≤ 20, both *CDKN3* and *UBE2C* had closer relationships with 5 genes; however, 5 genes did not appear in *EZH2*-based PPI network (Figure 1D). As we know, the cell cycle pathway is strongly associated with the occurrence of many types of tumors. Thus *CDKN3* and *UBE2C* genes were selected as key genes implicated in early-stage LUAD.

4.5. Survival Analysis of Key Gene

KM (LR test) method was used to evaluate the associations of the two key genes and OS. The results showed that low mRNA expression of *UBE2C* and *CDKN3* resulted in a higher OS rate than high mRNA expression (P = 0.037, 0.019, respectively) (Figure 2A). The mRNA expressions of *UBE2C* and *CDKN3* were significantly higher in early-stage LUAD tissues than that in normal tissues (P < 0.01) (Figure 2B).

The expression analysis based on GEPIA database showed that *UBE2C* and *CDKN3* were significantly highly expressed in all-stage LUAD tissues than in normal tissues (P < 0.01) (Figure 2C), and low mRNA expression of *UBE2C* and *CDKN3* resulted in a higher OS rate than high expression (P = 0.021, 0.00021, respectively) (Figure 2D). Gene expression analysis in various stages showed that the expressions of *UBE2C* and *CDKN3* were significantly different in four stages, and the expression of both genes was at the lowest level in early-stage LUAD tissues (P =

0.00245, 0.00398, respectively) (Figure 2E). Pearson correlation analysis showed that both genes had similar expression patterns in LUAD tissues (R = 0.61, P = 0) and normal tissues (R = 0.6, P = 4.5e-7) (Figure 3).

5. Discussion

LUAD is a complex malignant disease caused by gene aberration with a very poor 5-year OS. Identifying prognostic biomarkers in early-stage LUAD will contribute to offer a favorable prognosis. However, a widely accepted prognosticator has still not been found. A consistent reliable prognosticator based on gene expression is urgently required. In this study, we utilized bioinformatics strategy to integrate and analyze two early-stage LUAD-related gene expression profiles. Finally, we identified that *UBE2C* and *CDKN3* genes were significantly associated with the prognosis of early-stage LUAD patients, and low mRNA expressions of *UBE2C* and *CDKN3* resulted in a higher OS rate.

CDKN3 gene encodes the protein cyclin-dependent kinase inhibitor 3 that belongs to the dual specificity protein phosphatase family and possesses dual specificity phosphatase active toward substrates containing either phosphotyrosine or phosphoserine residues (22). The *CDKN3* plays important roles as oncogene or tumor suppressor gene in cell cycle regulation (23, 24). Many studies showed that CDKN3 was able to promote tumor development and progression in many tumors such as gastric cancer, breast cancer, cervical cancer, colorectal cancer, and ovarian cancer (25-28). In gastric cancer tissues, *CDKN3* was frequently up-regulated and related to poor outcome (28). In breast cancer and prostate cancer, high expression of *CDKN3* could promote cancer cell proliferation and phenotypic transformation (24, 29). In ovarian cancer, high expression of *CDKN3* enhanced cell invasion (25). In LUAD, Fan et al. found that overexpression and high expression of *CDKN3* were associated with a poor survival rate in patients (30).

Table 2. Top 20 Genes Obtained by Three Centrality Methods

Rank	Subgraph Centrality		Degree Centrality		Closeness Centrality	
	Gene	Subgraph	Gene	Degree	Gene	Closeness
1	*UBE2C*	90198016	*IL6*	54	*IL6*	0.076923
2	*CCNB1*	89298184	*TOP2A*	32	*EDN1*	0.076099
3	*TOP2A*	88912720	*EDN1*	32	*FOS*	0.075210
4	*NDC80*	86147896	*EZH2*	28	*SPP1*	0.075100
5	*RRM2*	86138048	*UBE2C*	26	*VWF*	0.074746
6	*PRC1*	86119328	*CCNB1*	26	*EGR1*	0.074611
7	*BUB1B*	86119328	*NDC80*	25	*EZH2*	0.074530
8	*CDKN3*	84999024	*FOS*	25	*TIMP3*	0.074530
9	*KIF11*	84680504	*CDKN3*	24	*CTGF*	0.074449
10	*ZWINT*	84680480	*CDC20*	24	*COL1A1*	0.074368
11	*TTK*	84680480	*VWF*	24	*CD36*	0.074021
12	*EZH2*	83511584	*RRM2*	23	*MMP1*	0.073941
13	*CDC20*	83265864	*PRC1*	23	*BMP2*	0.073941
14	*CEP55*	82918712	*BUB1B*	23	*MMP7*	0.073888
15	*PTTG1*	82562288	*CEP55*	23	*CAV1*	0.073888
16	*KIF4A*	81135504	*KIF11*	22	*COL1A2*	0.073862
17	*NUSAP1*	81135504	*ZWINT*	22	*CDKN3*	0.073729
18	*DLGAP5*	81135504	*TTK*	22	*UBE2C*	0.073677
19	*MELK*	81135504	*PTTG1*	22	*ID1*	0.073677
20	*TPX2*	81135480	*KIF4A*	21	*TEK*	0.073677

Currently, *CDKN3* has been recommended as a good candidate survival biomarker and potential therapeutic target in some cancers such as cervical cancer (26). The *CDKN3* has not been reported as a prognostic biomarker of early-stage LUAD. Our results confirmed that the expression of *CDKN3* was higher in early-stage LUAD tissues than that in normal tissues and low expression of *CDKN3* increased longer OS rate of LUAD patients, which indicated that *CDKN3* might serve as a good prognostic biomarker for early-stage LUAD.

UBE2C gene encodes UBE2C protein that belongs to the ubiquitin-conjugating enzyme family. Abnormal expression of *UBE2C* gene can lead to an increase of chromosomal instability, and promote the occurrence and development of a tumor (31). Many studies have proved the carcinogenic effect of UBE2C in cervix, thyroid, nasopharynx, mammary gland, and lung (31). Currently, *UBE2C* has been identified as a prognostic marker of breast cancer, gastric cancer, glioma, and bladder cancer (32-37). In lung cancer, few studies reported that deregulation of UBE2C could aggravate NSCLC progression by repressing autophagy (38). The present results showed that *UBE2C* was highly expressed in early-stage LUAD tissues, and low expression of *UBE2C* was associated with a longer OS in LUAD patients, which indi-

cated that *UBE2C* could also be as a prognostic predictor of early-stage LUAD.

The cell cycle is the basic process of cell life, and numerous studies have shown that the cell cycle played key roles in the formation of various malignant tumors (39). The present study demonstrated that the cell cycle was significantly enriched (P = 2.37e-2) with 5 genes, including *BUB1B*, *CCNB1*, *CDC20*, *PTTG1*, and *TTK*. These 5 genes and two key genes had closer interaction relationships with each other (Figure 1B and D), which indicated that *CDKN3* and *UBE2C* might play roles by interacting with those 5 genes within the cell cycle. The results further revealed that *CDKN3* and *UBE2C* might serve as potential biomarkers of LUAD diagnosis and prognosis and therapeutic targets of LUAD therapy at an early stage.

The strength of the current study was first to integrate early-stage LUAD-related gene expression profiles to identify key genes implicated in early-stage LUAD using bioinformatics methods, including DEGA, KEGG pathway analysis, PPI network, centrality analysis, and survival analysis. This study not only found DEGs in early-stage LUAD and elucidated the relationships among DEGs but also identified key genes associated with OS of early-stage LUAD patients.
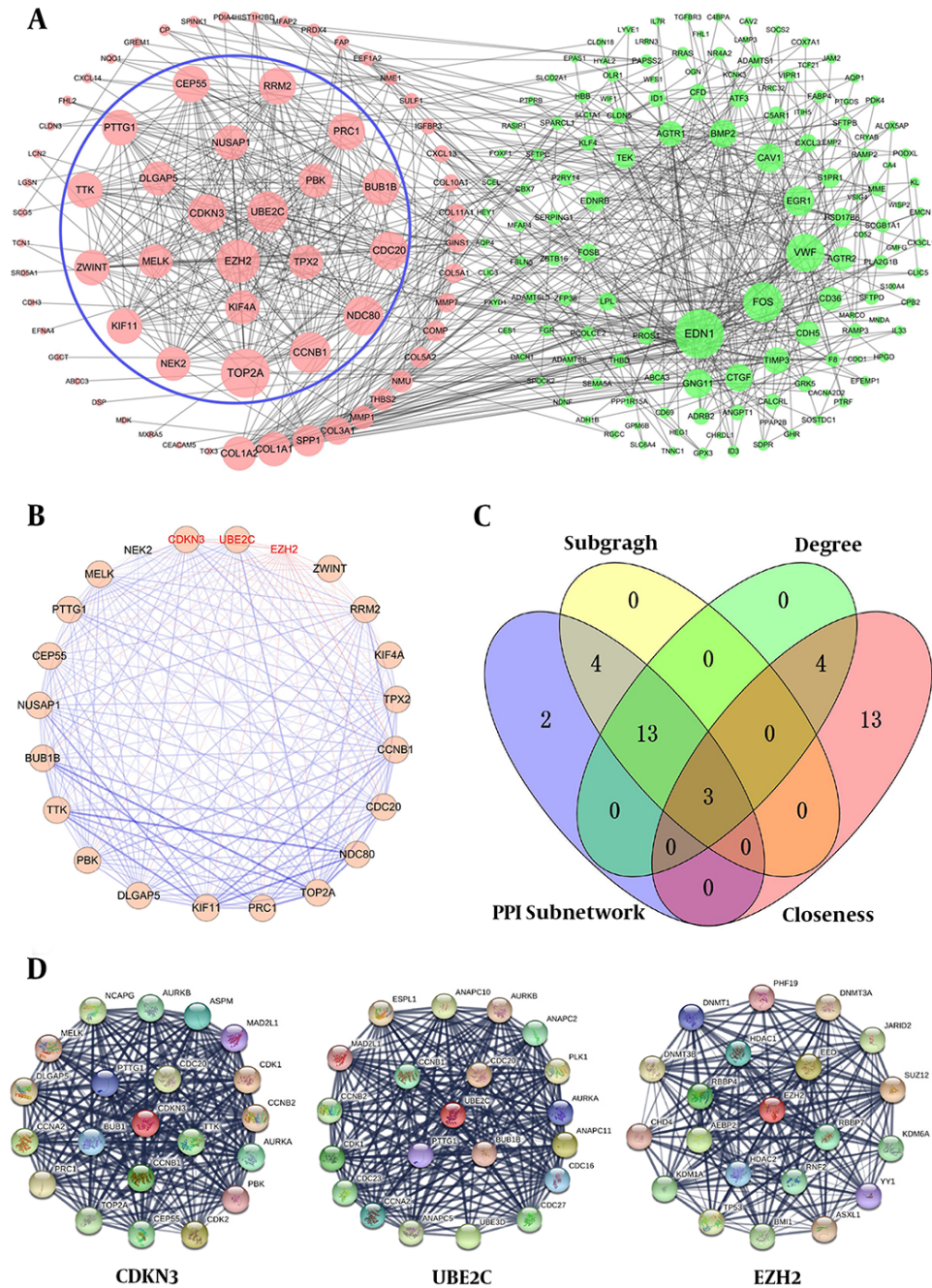
Figure 1. PPI network and module analysis. (A) Using STRING database, 207 (67 up-regulated and 140 down-regulated) of 303 DEGs were filtered into PPI network. The red and green nodes stood for up-regulated and down-regulated genes, respectively. Bigger nodes and labels represented genes with more links. PPI subnetwork in blue circle was the most significantly highly correlated module, and contained 22 nodes and 230 edges. The *CDKN3*, *UBE2C*, and *EZH2* genes were included in the PPI subnetwork and had more links. (B) The most significant PPI module consisted of 22 nodes with 230 edges. Bigger nodes represented genes with more links. Thicker edges represented higher combined scores among genes. Deeper color edges (red to blue) represented higher co-expression scores among genes. (C) Intersecting genes were identified by overlap analysis. Three genes, including *CDKN3*, *UBE2C*, and *EZH2* were identified as intersecting genes in early-stage lung adenocarcinoma. (D) PPI network of single gene based on STRING database.

The major limitation of the current study was that our results were obtained by pure bioinformatics methods. Although the results have been validated using public gene expression data, the results were not confirmed by experi-
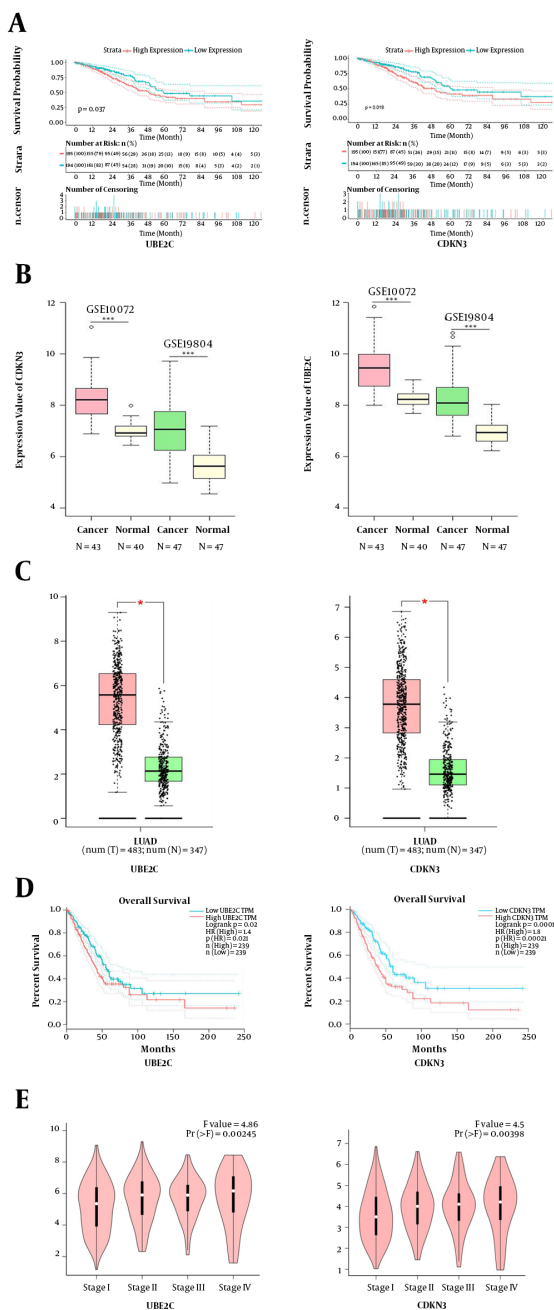
Figure 2. Survival curves and expression of key genes. (A) Low mRNA expression of *UBE2C* and *CDKN3* was significantly associated with overall survival of patients with early-stage LUAD. (B) The GSE10072 and GSE19804 data showed higher mRNA expression of *CDKN3* and *UBE2C* in early-stage LUAD tissues than in normal lung tissues (P < 0.01). (C) The expression analysis based on GEPIA database showed that *UBE2C* and *CDKN3* were significantly higher expression in all-stage LUAD tissues than in normal tissues. (D) Low mRNA expression of *UBE2C* and *CDKN3* resulted in a higher OS rate than high mRNA expression. (E) The mRNA expression of *UBE2C* and *CDKN3* was significantly different in four stages, and the mRNA expression of both genes was at the lowest level in early-stage LUAD tissues. LUAD: Lung adenocarcinoma.

ments. Next, we would verify the key genes associated with OS by experiments to confirm the association of key genes and OS.

5.1. Conclusion

In the present study, *UBE2C* and *CDKN3* were identified as key genes to play roles in early-stage LUAD and served as potential prognostic biomarkers for early-stage LUAD.
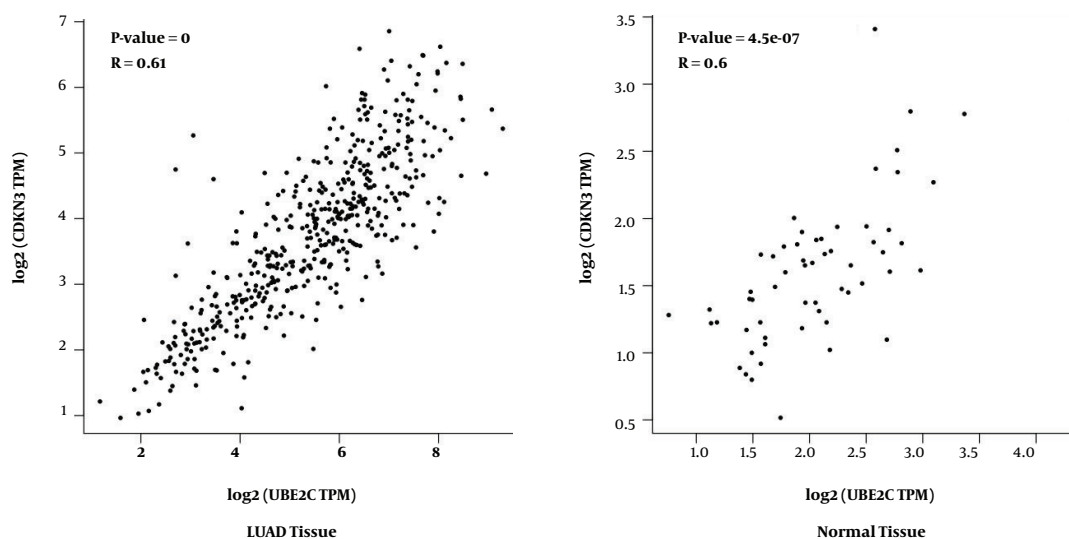
Figure 3. Correlation of *UBE2C* and *CDKN3* expression. In LUAD and normal tissues, Pearson correlation analysis showed that both genes had similar expression pattern. LUAD: Lung adenocarcinoma.

However, more experiments need to validate these prognosticators.

Acknowledgments

The authors thank Xiangqing Zhu and Xinghua Pan of Nation and Region Integrated Engineering Laboratory of Stem Cell and Immunocyte Biological Technology for helping to discuss this manuscript.

Footnotes

Authors' Contribution: Qiang Chen designed and supervised the study. Lutong Xu and Jing Hu searched the microarray from NCBI GEO database, performed DEGA, pathway enrichment analysis, PPI network construction, and survival analysis with the help of Tonglian Wang, Kang Zhang, and Yuanyue Li. Lutong Xu and Jing Hu contributed equally in this study. Qiang Chen wrote the manuscript with the help of Lutong Xu and Hongbo Zhao. Tao Shou revised the manuscript. All authors have read the manuscript and approved it.

Conflict of Interests: All authors declare that they have no competing interests.

Ethical Considerations: It is not declared by the author.

Financial Disclosure: All authors declare that they have no personal or professional financial conflicts.

Patient Consent: In this study, publicly available gene expression profiles from NCBI GEO database were collected with patients' consent approved by the relevant institutional review board.

References

1. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. *Adv Exp Med Biol*. 2016;**893**:1–19. doi: 10.1007/978-3-319-24223-1_1. [PubMed: 26667336].

2. Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C. Progress and prospects of early detection in lung cancer. *Open Biol*. 2017;**7**(9). doi: 10.1098/rsob.170070. [PubMed: 28878044]. [PubMed Central: PMC5627048].

3. Walters S, Maringe C, Coleman MP, Peake MD, Butler J, Young N, et al. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: A population-based study, 2004-2007. *Thorax*. 2013;**68**(6):551–64. doi: 10.1136/thoraxjnl-2012-202297. [PubMed: 23399908].

4. Wu K, House L, Liu W, Cho WC. Personalized targeted therapy for lung cancer. *Int J Mol Sci*. 2012;**13**(9):11471–96. doi: 10.3390/ijms130911471. [PubMed: 23109866]. [PubMed Central: PMC3472758].

5. Krzystanek M, Moldvay J, Szuts D, Szallasi Z, Eklund AC. A robust prognostic gene expression signature for early stage lung adenocarcinoma. *Biomark Res*. 2016;**4**:4. doi: 10.1186/s40364-016-0058-3. [PubMed: 26900477]. [PubMed Central: PMC4761211].

6. Nesbitt JC, Putnam JB Jr, Walsh GL, Roth JA, Mountain CF. Survival in early-stage non-small cell lung cancer. *Ann Thorac Surg*. 1995;**60**(2):466–72. [PubMed: 7646126].

7. Saito M, Shiraishi K, Kunitoh H, Takenoshita S, Yokota J, Kohno T. Gene aberrations for precision medicine against lung adenocarci-

noma. *Cancer Sci.* 2016;**107**(6):713–20. doi: 10.1111/cas.12941. [PubMed: 27027665]. [PubMed Central: PMC4968599].

8. Zhao J, Li L, Wang Q, Han H, Zhan Q, Xu M. CircRNA expression profile in early-stage lung adenocarcinoma patients. *Cell Physiol Biochem.* 2017;**44**(6):2138–46. doi: 10.1159/000485953. [PubMed: 29241190].

9. Pasche B, Grant SC. Non-small cell lung cancer and precision medicine: A model for the incorporation of genomic features into clinical trial design. *JAMA.* 2014;**311**(19):1975–6. doi: 10.1001/jama.2014.3742. [PubMed: 24846033].

10. He X, Zhang C, Shi C, Lu Q. Meta-analysis of mRNA expression profiles to identify differentially expressed genes in lung adenocarcinoma tissue from smokers and non-smokers. *Oncol Rep.* 2018;**39**(3):929–38. doi: 10.3892/or.2018.6197. [PubMed: 29328493].

11. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One.* 2008;**3**(2). e1651. doi: 10.1371/journal.pone.0001651. [PubMed: 18297132]. [PubMed Central: PMC2249927].

12. Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in non-smoking women. *Cancer Epidemiol Biomarkers Prev.* 2010;**19**(10):2590–7. doi: 10.1158/1055-9965.EPI-10-0332. [PubMed: 20802022].

13. Caliez J, Monnet I, Pujals A, Rousseau-Bussac G, Jabot L, Boudjemaa A, et al. [Lung adenocarcinoma with concomitant EGFR mutation and ALK rearrangement]. *Rev Mal Respir.* 2017;**34**(5):576–80. French. doi: 10.1016/j.rmr.2016.08.002. [PubMed: 27646667].

14. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004;**20**(3):307–15. doi: 10.1093/bioinformatics/btg405. [PubMed: 14960456].

15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;**43**(7). e47. doi: 10.1093/nar/gkv007. [PubMed: 25605792]. [PubMed Central: PMC4402510].

16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS.* 2012;**16**(5):284–7. doi: 10.1089/omi.2011.0118. [PubMed: 22455463]. [PubMed Central: PMC3339379].

17. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;**43**(Database issue):D447–52. doi: 10.1093/nar/gku1003. [PubMed: 25352553]. [PubMed Central: PMC4383874].

18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;**13**(11):2498–504. doi: 10.1101/gr.1239303. [PubMed: 14597658]. [PubMed Central: PMC403769].

19. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;**4**:2. [PubMed: 12525261]. [PubMed Central: PMC149346].

20. Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: A cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems.* 2015;**127**:67–72. doi: 10.1016/j.biosystems.2014.11.005. [PubMed: 25451770].

21. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017;**45**(W1):W98–W102. doi: 10.1093/nar/gkx247. [PubMed: 28407145]. [PubMed Central: PMC5570223].

22. Cress WD, Yu P, Wu J. Expression and alternative splicing of the cyclin-dependent kinase inhibitor-3 gene in human cancer. *Int J Biochem Cell Biol.* 2017;**91**(Pt B):98–101. doi: 10.1016/j.biocel.2017.05.013. [PubMed: 28504190]. [PubMed Central: PMC5641230].

23. Nalepa G, Barnholtz-Sloan J, Enzor R, Dey D, He Y, Gehlhausen JR, et al. The tumor suppressor CDKN3 controls mitosis. *J Cell Biol.* 2013;**201**(7):997–1012. doi: 10.1083/jcb.201205125. [PubMed: 23775190]. [PubMed Central: PMC3691455].

24. Yu C, Cao H, He X, Sun P, Feng Y, Chen L, et al. Cyclin-dependent kinase inhibitor 3 (CDKN3) plays a critical role in prostate cancer via regulating cell cycle and DNA replication signaling. *Biomed Pharmacother.* 2017;**96**:1109–18. doi: 10.1016/j.biopha.2017.11.112. [PubMed: 29196103].

25. Zhang LP, Li WJ, Zhu YF, Huang SY, Fang SY, Shen L, et al. CDKN3 knockdown reduces cell proliferation, invasion and promotes apoptosis in human ovarian cancer. *Int J Clin Exp Pathol.* 2015;**8**(5):4535–44. [PubMed: 26191143]. [PubMed Central: PMC4503015].

26. Barron EV, Roman-Bassaure E, Sanchez-Sandoval AL, Espinosa AM, Guardado-Estrada M, Medina I, et al. CDKN3 mRNA as a biomarker for survival and therapeutic target in cervical cancer. *PLoS One.* 2015;**10**(9). e0137397. doi: 10.1371/journal.pone.0137397. [PubMed: 26372210]. [PubMed Central: PMC4570808].

27. Yang C, Sun JJ. Mechanistic studies of cyclin-dependent kinase inhibitor 3 (CDKN3) in colorectal cancer. *Asian Pac J Cancer Prev.* 2015;**16**(3):965–70. [PubMed: 25735390].

28. Li Y, Ji S, Fu LY, Jiang T, Wu D, Meng FD. Knockdown of cyclin-dependent kinase inhibitor 3 inhibits proliferation and invasion in human gastric cancer cells. *Oncol Res.* 2017;**25**(5):721–31. doi: 10.3727/096504016X14772375848616. [PubMed: 27983933].

29. Deng M, Wang J, Chen Y, Zhang L, Xie G, Liu Q, et al. Silencing cyclin-dependent kinase inhibitor 3 inhibits the migration of breast cancer cell lines. *Mol Med Rep.* 2016;**14**(2):1523–30. doi: 10.3892/mmr.2016.5401. [PubMed: 27314680]. [PubMed Central: PMC4940103].

30. Fan C, Chen L, Huang Q, Shen T, Welsh EA, Teer JK, et al. Over-expression of major CDKN3 transcripts is associated with poor survival in lung adenocarcinoma. *Br J Cancer.* 2015;**113**(12):1735–43. doi: 10.1038/bjc.2015.378. [PubMed: 26554648]. [PubMed Central: PMC4701993].

31. Hao Z, Zhang H, Cowell J. Ubiquitin-conjugating enzyme UBE2C: Molecular biology, role in tumorigenesis, and potential as a biomarker. *Tumour Biol.* 2012;**33**(3):723–30. doi: 10.1007/s13277-011-0291-1. [PubMed: 22170434].

32. Loussouarn D, Campion L, Leclair F, Campone M, Charbonnel C, Ricolleau G, et al. Validation of UBE2C protein as a prognostic marker in node-positive breast cancer. *Br J Cancer.* 2009;**101**(1):166–73. doi: 10.1038/sj.bjc.6605122. [PubMed: 19513072]. [PubMed Central: PMC2713693].

33. Zhang HQ, Zhao G, Ke B, Ma G, Liu GL, Liang H, et al. Overexpression of UBE2C correlates with poor prognosis in gastric cancer patients. *Eur Rev Med Pharmacol Sci.* 2018;**22**(6):1665–71. doi: 10.26355/eurrev_201803_14578. [PubMed: 29630110].

34. Zhang J, Liu X, Yu G, Liu L, Wang J, Chen X, et al. UBE2C is a potential biomarker of intestinal-type gastric cancer with chromosomal instability. *Front Pharmacol.* 2018;**9**:847. doi: 10.3389/fphar.2018.00847. [PubMed: 30116193]. [PubMed Central: PMC6082955].

35. Psyrri A, Kalogeras KT, Kronenwett R, Wirtz RM, Batistatou A, Bournakis E, et al. Prognostic significance of UBE2C mRNA expression in high-risk early breast cancer. A hellenic cooperative oncology group (HeCOG) study. *Ann Oncol.* 2012;**23**(6):1422–7. doi: 10.1093/annonc/mdr527. [PubMed: 22056852].

36. Ma R, Kang X, Zhang G, Fang F, Du Y, Lv H. High expression of UBE2C is associated with the aggressive progression and poor outcome of malignant glioma. *Oncol Lett.* 2016;**11**(3):2300–4. doi: 10.3892/ol.2016.4171. [PubMed: 26998166]. [PubMed Central: PMC4774622].

37. Morikawa T, Kawai T, Abe H, Kume H, Homma Y, Fukayama M. UBE2C is a marker of unfavorable prognosis in bladder cancer after radical cystectomy. *Int J Clin Exp Pathol.* 2013;**6**(7):1367–74. [PubMed: 23826418]. [PubMed Central: PMC3693202].

38. Guo J, Wu Y, Du J, Yang L, Chen W, Gong K, et al. Deregulation of UBE2C-mediated autophagy repression aggravates NSCLC progression. *Oncogenesis.* 2018;**7**(6):49. doi: 10.1038/s41389-018-0054-6. [PubMed: 29904125]. [PubMed Central: PMC6002383].

39. Liu S, Yang TB, Nan YL, Li AH, Pan DX, Xu Y, et al. Genetic variants of cell cycle pathway genes predict disease-free survival of hepatocellular carcinoma. *Cancer Med.* 2017;**6**(7):1512–22. doi: 10.1002/cam4.1067.

[PubMed: 28639733]. [PubMed Central: PMC5504311].