Screening and identification of key biomarkers in lung squamous cell carcinoma by bioinformatics analysis

JUN MAN1, XIAOMEI ZHANG2, HUAN DONG1, SIMIN LI1, XIAOLIN YU1, LIHONG MENG1, XIAOFENG GU1, HONG YAN1, JINWEI CUI1 and YUXIN LAI1

1Department of Internal Medicine of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029; 2Department of Respiratory Medicine, Dongfang Hospital, Beijing University of Chinese Medicine, Beijing 100078, P.R. China

Received April 26, 2019; Accepted August 22, 2019

DOI: 10.3892/ol.2019.10873

Abstract. The high mortality rate of lung squamous cell carcinoma (LUSC) is in part due to the lack of early detection of its biomarkers. The identification of key molecules involved in LUSC is therefore required to improve clinical diagnosis and treatment outcomes. The present study used the microarray datasets GSE31552, GSE6044 and GSE12428 from the Gene Expression Omnibus database to identify differentially expressed genes (DEGs). Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were conducted to construct the protein-protein interaction network of DEGs and hub genes module using STRING and Cytoscape. The 67 DEGs identified consisted of 42 upregulated genes and 25 downregulated genes. The pathways predicted by KEGG and GO enrichment analyses of DEGs mainly included cell cycle, cell proliferation, glycolysis or gluconeogenesis, and tetrahydrofolate metabolic process. Further analysis of the University of California Santa Cruz and ONCOMINE databases identified 17 hub genes. Overall, the present study demonstrated hub genes that were closely associated with clinical tissue samples of LUSC, and identified TYMS, CCNB2 and RFC4 as potential novel biomarkers of LUSC. The findings of the present study contribute to an improved understanding of the molecular mechanisms of carcinogenesis and progression of LUSC, and assist with the identification of potential diagnostic and therapeutic targets of LUSC.

Introduction

Lung cancer is a disease with the highest morbidity and mortality rates worldwide. It is reported that non-small cell lung cancer (NSCLC) accounts for 85% of the total lung cancer cases worldwide, of which squamous cell lung cancer (LUSC), often with poor prognosis, accounted for 30% of NSCLC in 2017 (1,2). Data has demonstrated that more than 1 in 3 patients with lung adenocarcinoma (LUAD) benefit from molecular-targeted therapies (3). Inhibitors of epidermal growth factor receptor, v-ki-ras2 Kirsten rat sarcoma viral oncogene homologue and anaplastic lymphoma receptor tyrosine kinase are some of the few molecules that are targeted in lung cancer therapy (4). However, the application of molecular-targeted therapies in the diagnosis and treatment of LUSC in the clinical setting is very limited. Thus, the identification of biomarkers that are associated with the diagnosis and treatment of LUSC has become one of the main focus areas in research. An increasing number of studies have revealed new genetic changes associated with LUSC, including the onco-genes baculoviral IAP repeat contain 5 (BIRC5) and GAPDH. BIRC5 is an important inhibitor of apoptosis, which serves an important role in carcinogenesis and progression of LUSC (5). Li et al (6) reported a significantly higher expression level of BIRC5 in LUSC tissues compared with normal tissues, indicating the potential of BIRC5 as a target for anti-tumor therapy. On the other hand, GAPDH has been reported to serve a crucial role in regulating glycolysis in cancer cells. GAPDH depletes ATP in cancer cells via the inhibition of glycolysis, which eventually kills cancer cells (7,8). Hence, GAPDH has become a therapeutic target of interest against cancer cells. LUSC accounts for more than 400,000 deaths worldwide each year (2); it is important to highlight that the mortality rate of LUSC is inevitably high, even at early stage, despite several discoveries of potential targets such as BIRC5 and GAPDH. Thus, the investigation of other potential molecular mechanisms associated with LUSC is important.

In recent years, gene microarray and gene chip technologies have developed rapidly, which has provided a theoretical basis for the detection of genetic alterations in cancer cells (9,10). These technologies can be applied to identify differentially expressed genes (DEGs), which can potentially be associated with the carcinogenicity and progression of LUSC. In the
present study, in order to avoid false positive results from a single microarray gene expression dataset, three mRNA microarray datasets from the Gene Expression Omnibus (GEO) database platform were downloaded. LUSC tissues and non-cancerous tissues were analyzed in order to identify DEGs. Furthermore, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted and a protein-protein interaction (PPI) network was constructed in order to understand the molecular mechanisms underlying the generation and progression of LUSC. The associations between the hub genes and clinical tissue samples were identified using the University of California Santa Cruz (UCSC) Cancer Genomics Browser and ONCOMINE database. A total of 67 DEGs and 17 hub genes were identified as potential diagnostic and therapeutic biomarkers of LUSC. Five hub genes with the highest node value were selected via CentiScaPe, in which the results from the UCSC and ONCOMINE online clinical databases indicated all five hub genes to be associated with unfavorable prognosis of LUSC. Thymidylate synthetase (TYMS), cyclin B2 (CCNB2) and replication factor C subunit 4 (RFC4) were suggested as potential and novel target genes for the treatment of LUSC.

Materials and methods

Microarray data. GEO (http://www.ncbi.nlm.nih.gov/geo) is an open database of gene expression abundance, consisting of high throughout gene expression data, gene microarrays and gene chips. Three gene expression datasets [GSE31552 (11), GSE6044 (12) and GSE12428 (13)] were downloaded from the GEO (Affymetrix GPL6244 platform, Affymetrix Human Gene 1.0 ST Array; Affymetrix GPL201 platform, Affymetrix Human HG-Focus Target Array; Affymetrix GPL1708 platform, Agilent-012391 Whole Human Genome Oligo Microarray G4112A). The GSE31552 dataset included 25 LUSC tissue samples and 25 non-cancerous samples. GSE6044 included 15 LUSC tissue samples and 5 non-cancerous samples. GSE12428 included 34 LUSC tissue samples and 28 non-cancerous samples.

Identification of DEGs. The DEGs between LUSC and non-cancerous tissue samples were selected by GEO2R (http://www.ncbi.nlm.nih.gov/geo?geo2r). GEO2R is a GEO online analysis tool, which analyzes differential gene expression among two or more datasets in GEO. The P-values were adopted to screen the DEGs accurately, and the probe sets without corresponding gene symbols during the screening process were removed. LogFC (fold change) >1 or logFC<1 and P<0.05 were considered as statistically significant.

`KEGG pathway` and `Gene Oncology (GO)` enrichment analyses of DEGs. The Database for Annotation, Visualization and Integrated Discovery database (DAVID; http://david.ncifcrf.gov; version 6.8) is an online gene and pathway functional annotation database that contains biological information and also provides analysis tools (14). Biological information can be extracted from the comprehensive set of genes and proteins, which provides functional annotations. The KEGG database can be used to analyze genome information and gene function and study the gene expression information as a whole network (15). GO is a type of bioinformatics tool for annotating genes and analyzing their biological processes. GO enrichment analyses contain three modules of molecular function, cell composition and biological process (16). In order to analyze the function and cell signaling pathways of DEGs, KEGG and GO enrichment analyses were conducted using the DAVID database. KEGG and GO enrichment bubble plots were drawn using online graphics tools Image GP (http://www.ebihi.com/ImageGP/). P<0.05 was considered as statistically significant.

PPI network and hub gene module construction. The PPI network of DEGs was constructed by the online analysis website Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org; version 11.0) (17) and the interaction of a combined score >0.4 was considered as statistically significant. Analyzing the function of PPI can provide insights into the mechanisms of disease occurrence and development. Cytoscape (version 3.6.1) is an open bioinformatics software platform that can be used to construct a visual network of molecular interactions (18). The plug-in Molecular Complex Detection (MCODE) (version 1.4.2) of Cytoscape is an APP for detecting densely correlated regions in the PPI networks (19). The gene modules were visualized and graphically displayed with the plug-in MCODE. The selection criteria were as follows: MCODE score >5; node score cut-off, 0.2; degree cut-off, 2; k-score, 2; and Max depth, 100. CentiScaPe (version 2.2), a Cytoscape APP specifically designed to calculate centrality indexes for the selection of the most critical nodes in a network (20). The plug-in CentiScaPe 2.2 was used to identify hub genes for functional analysis with interaction node degrees ≥10.

Hub genes selection and analysis. The hub genes with interaction node degrees ≥10 were screened. The network of the genes and their co-expression genes was constructed by the online platform cBioPortal (http://www.cbioportal.org) (21,22). Hierarchical clustering of hub genes was constructed by online analysis website UCSC Cancer Genomics Browser (http://genome-cancer.ucsc.edu) (23). Heat maps of hub genes expression in three different studies of clinical LUSC

![Figure 1. Venn diagram. Genes with |logFC|>1 and P<0.05, among the mRNA expression profiling datasets GSE31552, GSE6044 and GSE12428, were selected as DEGs. The overlap among the three datasets resulted in 67 DEGs.](http://genome-cancer.ucsc.edu)
samples vs. non-cancerous tissue samples (24-26), and the associations between the expression patterns and tumor stage, overall survival status (the survival rate of patients from diagnosis to the end of the study) and survival status at 5 years (the survival rate five years after diagnosis) were analyzed using the Hou Lung dataset (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19188) (27), which was obtained from the Oncomine database (http://www.oncomine.com) (28,29).

Results

Identification of DEGs in LUSC. The standard microarrays were obtained from the GEO database platform. Following further analysis using GEO2R, DEGs were identified from the GSE31552 (957), GSE6044 (724) and GSE12428 (745) datasets. The 67 DEGs between the three datasets are presented in a Venn diagram (Fig. 1), consisting of 42 upregulated genes and 25 downregulated genes between LUSC and non-cancerous tissues.

KEGG and GO enrichment analyses of DEGs. Functional and pathway enrichment analyses of DEGs were conducted by DAVID to obtain the biological classification. The GO enrichment analysis included biological processes (BP), cell component (CC) and molecular function (MF) terms of the DEGs. The results of the KEGG and GO enrichment analyses are presented as bubble plots in Fig. 2. Changes in BP were significantly enriched in ‘cell division’, ‘positive regulation of cell proliferation’, ‘negative regulation of endopeptidase activity’ and ‘tetrahydrofolate metabolic process’ (Fig. 2A). Changes in CC were significantly enriched in ‘extracellular exosome’, ‘extracellular space’, ‘cytosol’ and ‘vesicle’ (Fig. 2B). Changes in MF were mainly enriched in ‘protein binding’, ‘cysteine-type endopeptidase inhibitor activity’, ‘protein binding’ and ‘identical protein binding’ (Fig. 2C). The KEGG pathway analysis was mainly enriched in ‘cell cycle’, ‘glycolysis or gluconeogenesis’, ‘metabolic pathways’ and ‘one carbon pool by folate’ (Fig. 2D).

PPI network and hub gene module construction. In order to identify the hub genes of LUSC, the PPI network of DEGs was analyzed by STRING. The results revealed that most genes interacted with each other and were located in the center of the network, and were closely associated with the surrounding proteins in the network (Fig. 3A). To enhance the accuracy of the results, the PPI network was also analyzed by Cytoscape. The obtained results were in correspondence with the results of STRING (Fig. 3B), and the hub gene module was obtained using MCODE (Fig. 3C).

Hub gene selection and analysis. In total, 17 genes were regarded as hub genes with degrees ≥10 using CentiScaPe. The names, abbreviations and functions for each of these hub genes

Figure 2. KEGG and GO enrichment plots of DEGs. The plots from the GO enrichment analysis of DEGs for (A) biological processes (B) cell and (C) molecular function were obtained using DAVID and drawn using Image GP. (D) KEGG enrichment analysis was performed using DAVID, the bubble plot of KEGG analysis was drawn using Image GP. KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; DAVID, Database for Annotation, Visualization and Integrated Discovery.
Figure 3. PPI network and hub gene module of DEGs. (A) The PPI network was established with the Search Tool for the Retrieval of Interacting Gene tool. (B) The PPI network of DEGs was established using Cytoscape. (C) The hub gene module was generated from the PPI network, consisting of 15 nodes and 84 edges. Upregulated genes are marked in red. Downregulated genes are marked in blue. DEGs, differentially expressed genes; PPI, protein-protein interaction.
are summarized in Table I. The genes associated with the hub genes and their co-expression network were obtained using the cBioPortal online platform by performing interaction analysis (Fig. 4A). The expression of 17 hub genes in LUSC tissues and its association with the severity and prognosis among LUSC patients were further explored using the UCSC and ONCOMINE online databases. Furthermore, a heat map of hierarchical clustering obtained using UCSC demonstrated that the expression of hub genes in LUSC tissues was higher compared with that of non-cancerous samples. However, the expression of hub genes showed no differences with gender (Fig. 4B). The heat map of hub genes expression in clinical LUSC tissue samples and normal tissue samples were analyzed using three different datasets, using the ONCOMINE online platform. The results revealed that most of the hub genes were significantly upregulated in clinical LUSC samples in all the datasets (Fig. 5). The hub genes whose interaction node degree was among the top five were TYMS, CCNB2, RFC4, BIRC5 and GAPDH, indicating their potential role in the processes of carcinogenesis, development and unfavorable prognosis of LUSC. The associations between the upregulated hub genes and tumor stage, overall survival status and survival status at 5 years were also analyzed. The top five genes were associated with high tumor stage, poor overall survival status and poor survival status at 5 years, which suggests their upregulation to be involved in the promotion of tumor progression and poor prognosis (Fig. 6).

### Discussion

In recent years, the incidence and mortality of lung cancer has continued to increase rapidly worldwide (30). LUAD and LUSC are the two common types of lung cancer. The incidence of LUSC was reported to be associated with smoking, whereas the treatment of LUSC remains limited compared with that of LUAD. The underlying pathological mechanisms of LUSC at the molecular level are still at the exploration stage (31). Mutations or amplifications of phosphatidylinositol-3

| No. | Gene symbol | Full name                             | Function                                                                 |
|-----|-------------|---------------------------------------|--------------------------------------------------------------------------|
| 1   | TYMS        | Thymidylate synthetase                | DNA synthesis, DNA repair and proliferation of cancer cells.            |
| 2   | CCNB2       | Cyclin B2                             | Correlated with invasion, metastasis and poor prognosis of various cancer types. |
| 3   | RFC4        | Replication factor C subunit 4        | Associated with poorly differentiated and advanced Tumor-Node-Metastasis stage in multiple cancer types. |
| 4   | BIRC5       | Baculoviral IAP repeat containing 5    | Plays a key role in proliferation, apoptosis and angiogenesis of LUSC.   |
| 5   | GAPDH       | Glyceraldehyde-3-Phosphate dehydrogenase | Elevated GAPDH significantly promotes cell proliferation and migration in LUSC. |
| 6   | CKS1B       | CDC28 protein kinase regulatory subunit 1B | Attributes to prognosis, chemo resistance and chemosensitivity in cancer. |
| 7   | MCM6        | Minichromosome maintenance complex component 6 | Initiation of DNA replication and a marker for proliferating cells.     |
| 8   | EZH2        | Enhancer of Zeste 2 polycomb repressive complex 2 subunit | Invasion, cell proliferation and adverse prognosis in LUSC.            |
| 9   | PTTG1       | Pituitary tumor-transforming 1         | Carcinogenesis, migration, invasion and prognosis.                      |
| 10  | CDK4        | Cyclin dependent kinase 4             | Regulates cell cycle positively, overexpressed and gene amplified in LUSC. |
| 11  | TPX2        | TPX2, microtubule nucleation factor   | Overexpression associated with differentiation grade, stage and metastasis of LUSC. |
| 12  | PRG1        | Protein regulator of cytokinesis 1     | Promotes progression and migration in LUSC.                             |
| 13  | CKS2        | CDC28 protein kinase regulatory subunit 2 | Inhibit DNA damage response and contribute to tumor cell proliferation in breast cancer |
| 14  | CDC45       | Cell division cycle 45                | Regulator of cell proliferation and associated with S-phase DNA damage. |
| 15  | KPNA2       | Karyopherin subunit alpha 2           | Attributed to cancer cell proliferation and metastasis.                |
| 16  | NCAPG       | Non-SMC condensin I complex subunit G  | Associated with cell cycle, apoptosis and migration in human hepatocellular carcinoma. |
| 17  | UBE2S       | Ubiquitin conjugating enzyme E2 S      | Involved in the malignant characteristics, mitosis and survival of various types of cancer cells. |
kinases (PI3K), phosphatase and Tensin homolog (PTEN), erythropoietin-producing hepatocellular A2 (EphA2) and liver kinase B1 (LKB1) were reported to be associated with the incidence, progression and prognosis of LUSC (32,33). A study conducted using the Cancer Genome Atlas Research Network demonstrated the dysfunction of NFE2L2, KEAP1, CDKN2A and RB1, and the abnormal structures of their products are associated with the occurrence and development of LUSC (34). The high mortality rate of LUSC is in part due to the lack of early detection of LUSC biomarkers (35). As a result, the identification of key molecules involved in LUSC is required and important for improving clinical efficacy. Microarray is a high-throughput technology in obtaining novel biomarkers, which can provide the basis for further studies on the mechanism of LUSC and clinical targeted therapies at the molecular level.

In the present study, three mRNA microarray datasets were analyzed to identify 67 common DEGs. The DEGs consisted of 42 upregulated and 25 downregulated genes between LUSC tissue samples and normal tissue samples. GO terms and KEGG pathway enrichments were analyzed in order to investigate interactions among the DEGs. The results indicated that
Figure 5. ONCOMINE analysis of LUSC vs. normal tissue of hub genes. Heat maps of hub genes expression in clinical LUSC samples vs. normal tissue samples. Hub genes: (A) TYMS, (B) CCNB2, (C) RFC4, (D) BIRC5, (E) GAPDH, (F) CkS1B, (G) MCM6, (H) EZH2, (I) PTTG1, (J) CDk4, (K) TPX2, (L) PRC1, (M) CKS2, (N) CDC45, (O) KPNA2, (P) NCAPO and (Q) UBE2S. The rank for a gene is the median rank for that gene across each of the analyses and the P-value for a gene is its P-value for the median-ranked analysis. LUSC, lung squamous cell carcinoma.
Figure 6. The associations between upregulation of TYMS, CCNB2, RFC4, BIRC5 and GAPDH and tumor stage, overall survival status and survival status at 5 years in the Hou Lung dataset which was obtained from the ONCOMINE database. (A–C) TYMS, (D–F) CCNB2, (G–I) RFC4, (J–L) BIRC5 and (M–O) GAPDH mRNA expression in LUSC vs. normal lung tissues. LUSC lung squamous cell carcinoma.
the 67 DEGs were significantly enriched in cell cycle, cell proliferation, glycolysis or gluconeogenesis, and tetrahydrofolate metabolic process. Previous studies have illustrated that dysregulations of cell cycle and cell proliferation serve roles in the carcinogenesis and malignant change of LUSC (36-38). In addition, multiple studies have also shown that glycolysis or gluconeogenesis serve important roles in tumor initiation, progression and unfavorable prognosis in cancer (39,40). Furthermore, gene polymorphism of tetrahydrofolate induces a decreased activity of tetrahydrofolate reductase, which affects the normal metabolism of folate in cells, where tetrahydrofolate metabolic disorder is closely associated with tumorigenesis (41,42). The findings of the present study were in accordance with the conclusions of previous studies and showed that GO and KEGG enrichment analyses were significantly enriched in cell cycle, cell proliferation, glycolysis or gluconeogenesis, and tetrahydrofolate metabolic process. A total of 17 DEGs were identified as hub genes with an interaction node degree ≥10. The hub genes whose degrees were among the top five were TYMS, CCNB2, RFC4, BIRC5 and GAPDH, and the PPI network showed that they were directly interacting with each other.

Several studies have suggested that TYMS is a predictive biomarker to test for the effectiveness of pemetrexed used in chemotherapy for treating NSCLC (43,44). Lu et al (45) reported that the expression of TYMS was significantly upregulated among patients with lymph node metastasis. The expression of TYMS was also higher among patients with 5-year recurrence rate. Besides, high expression of TYMS was found in the case of breast cancer, which resulted in increased susceptibility of an individual to the progression of the disease. Gene polymorphisms of TYMS have been reported to have potential in improving the diagnosis, prevention and treatment of breast cancer (46). Hence, several studies are now focusing on the association between gene polymorphic variations of TYMS with various types of cancer (47-49). BIRC5 was upregulated in 76% LUSC samples and the expression in LUSC tissues was significantly higher compared with that in non-cancerous tissues (50). BIRC5 is a potential biomarker or therapeutic target of smoking-associated LUSC (51). The expression of BIRC5 was higher among patients who are smokers compared with non-smokers, and in squamous vs. non-squamous lung tumor (P<0.001). The present study also demonstrated that BIRC5 expression level was negatively associated with the expression of tumor suppressor gene Tp53 (52). GAPDH serves a critical role in inhibiting the process of glycolysis in tumor cells (53). The expression of GAPDH was notably upregulated in LUSC tissues, and an increased level of GAPDH significantly promotes the cell proliferation and migration in LUSC (54). CCNB2 as a member of the cell cyclin protein family, was significantly associated with different staging and metastatic statuses of tumors (P<0.001). Thus, CCNB2 is a potential biomarker for evaluating metastatic status and therapeutic efficacy for cancer patients (55). Additionally, the upregulation of CCNB2 was closely associated with the degree of differentiation, progression, lymph node metastasis, invasion and adverse prognosis in NSCLC (56). CCNB2 has also been found to be upregulated in patients with bladder and colorectal cancer (57). Thus, CCNB2 is a potential diagnostic biomarker and a therapeutic target for LUSC. RFC4 was involved in DNA replication and regulation of cell proliferation and cell cycle. Studies reported an association of RFC4 with cancer progression and worse survival outcome, and the ability to predict response to radiotherapy and neoadjuvant radiotherapy in rectal cancer (58,59). RFC4 has been demonstrated to be associated with several types of cancer, however the underlying carcinogenic mechanism needs to be further explored. The top five hub genes reported in the present study were associated with various types of cancer. Multiple studies have demonstrated that GAPDH and BIRC5 are associated with LUSC (50,54). However, to our knowledge, no previous studies reported TYMS, CCNB2 and RFC4 to be directly associated with LUSC. In the present study, ONCOMINE and UCSC analysis confirmed that the top five hub genes from clinical LUSC samples were significantly upregulated and were all associated with different staging of cancer and survival rate compared to that of other samples. Therefore, TYMS, CCNB2 and RFC4 are potential novel biomarkers of LUSC for further investigation.

Among the other 12 hub genes identified in the present study, MCM6, EZH2, CDK4, TPX2 and PRC1 were previously reported to be associated with LUSC. Minichromosome maintenance (MCM) proteins serve a critical role in cell proliferation and cell cycle. Meanwhile, MCM6 is often associated with poor prognosis, particularly among male patients with LUSC and with a history of smoking (60). The presence of EZH2 was associated with the aggressiveness of cancer development. Behrens et al (61) analyzed 221 LUSC samples and 320 lung adenocarcinomas samples, which revealed significantly higher expression of EZH2 in LUSC compared with that of LUAD (P<0.0001). Cell cycle protein CDK4 has an established association with neoplasia and cancer progression. Recent studies have found that pathways including that of CDK4/6 were frequently altered in LUSC via diverse mechanisms, suggesting CDK4/6 inhibitors as potential target for the treatment of LUSC (62,63). TPX2, which actively participates in the formation of spindle microtubules during mitosis, was significantly associated with cell differentiation and metastatic status of LUSC cells, suggesting its potential as a prognostic predictor of LUSC (64). PRC1, an important protein involved in cytokinesis, plays an important role in microtubule organization in eukaryotes (65). A recent study has implicated the overexpression of PRC1 in LUSC to be associated with increased susceptibility to lymph node metastasis and shorter survival time in patients with LUSC (66).

Literature review revealed that the interaction among LUSC and hub genes CKS1B, PTTG1, CKS2, CDC45, KPNA2, NCPAG and UBE2S has not been widely reported. These genes are potential novel biomarkers and therapeutic targets of LUSC. CKS1B, an adaptor for cyclin-dependent kinases, was shown to be associated with chemoresistance, low chemosensitivity and poor prognosis in cancer. An elevated level of CKS1B has been reported to result in the resistance of cancer cells to bortezomib, and activation of the NEDD8 pathway, which in turn leads to further advancement of cancer. Thus, CKS1B is a potential novel target in multiple myeloma (67). In addition, CKS1B promotes proteasomal degradation and ubiquitination of p27Kip1. Overexpression of CKS1B contributes to an increased turnover rate of p27Kip1 and promotion of cancer cell proliferation, resulting in poor
prognosis in many types of cancer (68). PTTG1 functions to regulate transcription, the G-M phase of mitosis and the repair of DNA. Overexpression of PTTG1 has been reported in multiple types of cancer, in which PTTG1 is associated with metabolic processes, such as carcinogenesis, migration, invasion and epithelial-mesenchymal transition in squamous cell carcinomas (69). Recent studies demonstrated that various non-coding RNAs lead to cancer cell growth and metastasis via PTTG1 (70,71). Therefore, PTTG1 is a potential and novel therapeutic target of LUSC tumor growth and metastasis. CKS2, as a cyclin-dependent kinase-interacting protein, serves important roles in regulating cell cycle, inducing apoptosis, as well as regulating cancer cell invasion and metastasis. Upregulation of CKS2 can lead to DNA damage in cells, which can increase the proliferation of cancer cell (72). Therefore, the molecular mechanism by which CKS2 regulates cell cycle and induces cell apoptosis may be critical in investigating diagnostic methods and treatment methods for LUSC. CDC45 is an essential regulator of cell proliferation. An elevated expression of CDC45 is correlated with DNA damage in the S-phase, in which the anti-cancer effect of CDC45 suppressor is mediated by limiting DNA damage during S phase (73). KPN2, as a member of the Karyopherin α family, actively participates in the process of signal transduction from the extracellular space to the nucleus. Furthermore, the upregulation of KPN2 is correlated with cancer cell proliferation and metastasis (74). Moreover, several studies reported that KPN2 is involved in the progression of cancer by regulating nuclear translocation of cancer-associated proteins; which may explain the significantly upregulated expression of KPN2 in LUSC. NCAPG is a novel mitotic gene and provides novel therapeutic targets for cancer. Goto et al (75) reported that NCAPG, as a target of miR-145-3p, could predict the survival rate of patients with prostate cancer. UBE2S is a central protein in the process of the miR-145-3p, could predict the survival rate of patients with cancer. UBE2S is a novel mitotic gene and provides novel therapeutic targets for cancer. Goto et al (75) reported that NCAPG, as a target of miR-145-3p, could predict the survival rate of patients with prostate cancer. UBE2S is a central protein in the process of ubiquitination and is associated with the malignancy of various types of cancer (76). Studies have reported the upregulation of UBE2S to enhance the nuclear translocation of β-catenin and induced expression of c-Myc and cyclin D1, suggesting UBE2S as a potential prognostic factor and oncogene in LUSC (77-78).

In conclusion, the present study was conducted in order to identify potential DEGs that may be associated with carcinogenesis or adverse progression of LUSC. A total of 67 DEGs and 17 hub genes were identified, in which hub genes were regarded as promising targets for the diagnosis and treatment of LUSC. Meanwhile, TYMS, CCNB2 and RFC4 were identified as potential novel biomarkers of LUSC. However, the present study was based on bioinformatics methods and no experiments were performed to validate the findings. Therefore, further studies are required to explore the biological association between the genes identified in the present study in LUSC, in order to improve treatments and clinical outcomes of LUSC.

Acknowledgements

Not applicable.

Funding

The present study was financially supported by the National Natural Science Foundation of China (grant no. 81273696).

Availability of data and materials

The datasets used and/or analyzed in the present study are available from the corresponding authors on reasonable request.

Authors’ contributions

JM, XZ, HD and SL designed the study. JM and XZ analyzed and interpreted the microarray datasets, and wrote the manuscript. XY and LM made substantial contributions to acquisition and interpretation of data. JM, XG, HY, JC and YL analyzed the data. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Youlデン DR, Cram SM and Baade PD: The International Epidemiology of Lung Cancer: Geographical distribution and secular trends. J Thorac Oncol 3: 819-831, 2008.
2. Lazarus KA, Hadi F, Zambon E, Bach K, Santolla MF, Watson JK, Correia LL, Das M, Ugur R, Pesna S, et al: BCL11A interacts with SOX2 to control the expression of epigenetic regulators in lung squamous carcinoma. Nat Commun 9: 3327, 2018.
3. Kris MG, Johnson BE, Kwiatkowski DJ, Iafate AJ, Wistuba II, Aronson SL, Engelmann JA, Shyr Y, Khuri FR, Rudin CM, et al: Identification of driver mutations in tumor specimens from 1,000 patients with lung adenocarcinoma: The NCI’s Lung Cancer Mutation Consortium (LCMC). J Clin Oncol 29: CRA7506, 2011.
4. Chan BA and Hughes BG: Targeted therapy for non-small cell lung cancer: Current standards and the promise of the future. Transl Lung Cancer Res 4: 36-54, 2015.
5. Vayshilya NA, Zinovyeva MV, Sass AV, Kopantzev EP, Vinogradova TV and Sverdlov ED: Increased expression of BIRC5 in non-small cell lung cancer and esophageal squamous cell carcinoma does not correlate with the expression of its inhibitors SMAC/DIABLO and PML. Mol Biol 42: 579-587, 2008.
6. Li S, Sun X, Miao S, Liu J and Jiao W: Differential protein-coding gene and long noncoding RNA expression in smoking-related lung squamous cell carcinoma. Thorac Cancer 8: 672-681, 2017.
7. Ganapathy-Kanniappan S, Geschwind JF, Kunjithapatham R, Buijs M, Vossen JA, Tchernyshyov I, Cole RN, Syed LH, Rao PP, Ota S and Vali M: Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is pyruvylated during 3-bromopyruvate mediated cancer cell death. Anticancer Res 29: 4909-4918, 2009.
8. Xu RH, Pelicano H, Zhou Y, Carew JS, Feng L, Bhalla KN, Keating MJ and Huang P: Inhibition of glycolysis in cancer cells: A novel strategy to overcome drug resistance associated with mitochondrial respiratory defect and hypoxia. Cancer Res 65: 613-621, 2005.
9. Dawany NB and Tozeren A: Asymmetric microarray data produces gene lists highly predictive of research literature on multiple cancer types. BMC Bioinformatics 11: 483, 2010.
10. Dawany NB, Dampier WN and Tozeren A: Large-scale integration of microarray data reveals genes and pathways common to multiple cancer types. Int J Cancer 128: 2881-2891, 2011.
11. Lin J, Marquardt G, Mullapudi N, Wang T, Han W, Shi M, Keller S, Zhu C, Locker J and Spivack SD: Lung cancer transcriptomes refined with laser capture microdissection. Am J Pathol 184:2428-2434, 2014.

12. Rhoerbeck A, Neukirchen J, Rosskopf M, Pardillos GS, Geddert H, Schwalen A, Gabbert HE, von Haeseler A, Pitschke G, Schott M, et al: Gene expression profiling for molecular distinction and characterization of laser captured primary lung cancers. J Transl Med 6:69, 2008.

13. Boelens MC, van den Berg A, Fehrmann RS, Geelings M, de Jong WK, te Meerman GJ, Sietsma H, Timens W, Postma DS and Groen HJ: Current smoking-specific gene expression signature in normal bronchial epithelium is enhanced in squamous cell lung cancer. J Pathol 218:182-191, 2009.

14. Harte DW, Tan Q, Keel C, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC and Lempicki RA: DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res 35:W169-W175, 2007.

15. Kanekaris M, Furumichi M, Tanabe M, Sato Y and Morishima K: KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353-D361, 2017.

16. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25-29, 2001.

17. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al: STRING v10: Protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447-D452, 2015.

18. Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27:431-432, 2011.

19. Bader GD and Hogue CW: An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4:2, 2003.

20. Scardoni G, Tosadori G, Faizan M, Spoto F, Fabbrini F and Laudanna C: Biological network analysis with CentiScaPe: Centralities and experimental dataset integration. F1000Res 3:25, 2014.

21. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross BE, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al: Integrative analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics 21:4205-4208, 2005.

22. Lee CH, Lee MK, Kang CD, Kim YD, Park DY, Kim JY, Sol MY and Suh KS: Differential expression of hypoxia inducible factor-1 alpha and tumor cell proliferation between squamous cell carcinomas and adenocarcinomas among operable NSCLC cell lung carcinomas. J Korean Med Sci 18:196-203, 2003.

23. Hirano T, Franzen B, Kato H, Ebihara Y and Auer G: Genesis of squamous cell lung carcinoma. Sequential changes of proliferation, DNA ploidy, and p53 expression. Am J Pathol 144:296-302, 1994.

24. Liu GM and Zhang YM: Targeting FBPase is an emerging novel approach for cancer therapy. Cancer Cell Int 18:36, 2018.

25. Hanahan D and Weinberg RA: Hallmarks of cancer: The next generation. Cell 144:646-674, 2011.

26. Xiao P, Keku T, Millikan R, Worley K, Winkel S, Eaton A, Biscocho L, Chao YL and Anders CK: TYMS gene polymorphisms in breast and prostate cancer patients receiving 5-fluorouracil-based chemotherapy. J Cancer Res Treat 49:1478-1483, 2013.

27. Fuji T, Toyooka S, Ichimura K, Fujisawa Y, Hotta K, Soh J, Suhelisa H, Kobayashi N, Aoe M, Yoshino T, et al: ERCC1 protein expression predicts the response of cisplatin-based neoadjuvant chemotherapy in non-small-cell lung cancer. Lung Cancer 59:377-384, 2010.

28. Casai D, Ozasa H, Oguri T, Miyazaki M, Uemura T, Takakowa U, Kumi E, Ohkubo H, Maeno K and Niimi A: Thymidylate synthase gene copy number as a predictive marker for response to pemetrexed treatment of lung adenocarcinoma. Anticancer Res 33:1935-1940, 2013.

29. Lu Y, Zhuo C, Cui B, Liu Z, Zhou P, Lu Y and Wang B: TYMS serves as a prognostic indicator to predict the lymph node metastasis in Chinese patients with colorectal cancer. Clin Biochem 46:1478-1483, 2013.

30. He Y, Penney ME, Negandhi AA, Parfrey PS, Savas S and Yilmaz YE: XRCC3 Thr241Met and TYMS variable number tandem repeat polymorphisms are associated with time-to-metastasis in colorectal cancer. PLoS One 13:e0192316, 2018.
50. Knizhnik AV, Kovaleva OB, Laktonin KK, Mochal'tnikova VV, Komelev AV, Chevkina EM and Zborovskaya IB: Arf6, RalA and BirC5 protein expression in non small cell lung cancer. Mol Biol (Mosk) 45: 307-315, 2011 (In Russian).

51. Yu X, Zhang Y, Cavaezos D, Ma X, Zhao Z, Du L and Pertsemidis A: miR-195 targets cyclin D3 and survivin to modulate the tumorigenesis of non-small cell lung cancer. Cell Death Dis 9: 193, 2018.

52. Phiboonchaiyanan PP, Petsirpen N, Sritularak B and Chamvorachote P: Phoyunmanin E induces apoptosis of non-small cell lung cancer cells via p53 activation and down-regulation of survivin. Anticancer Res 38: 6281-6290, 2018.

53. Ganapathy-Kannianpan S: Evolution of GAPDH as a drug-gable target of tumor glycolysis? Expert Opin Ther Targets 22: 295-298, 2018.

54. Hao L, Zhou X, Liu S, Sun M, Song Y, Du S, Sun B, Guo C, Gong L, Hu J, et al: Elevated GAPDH expression is associated with the proliferation and invasion of lung and esophageal squamous cell carcinomas. Proteomics 15: 3087-3100, 2015.

55. Mo ML, Chen Z, Li J, Li HL, Sheng Q, Ma HY, Zhang FX, Huo YW, Zhang X, Sun DQ, et al: Use of serum circulating CCNB2 in cancer surveillance. Int J Biol Markers 25, 2010.

56. Takashima S, Saito H, Takahashi N, Imai K, Kudo S, Atarai M, Saito Y, Motoyama S and Minamiya Y: Strong expression of cyclin B2 mRNA correlates with a poor prognosis in patients with non-small cell lung cancer. Tumour Biol 35: 4257-4265, 2014.

57. Lei CY, Wang W, Zhu YT, Fang WY and Tan WL: The decrease of cyclin B2 expression inhibits invasion and metastasis of bladder cancer. Urol Oncol 34: 237.e1-e10, 2016.

58. Wang XC, Yue X, Zhang RX, Liu TY, Pan ZZ, Yang MI, Lu ZH, Wang ZY, Peng JH, Le LY, et al: Genome-wide RNAi screening identifies RFC4 as a factor that mediates radioresistance in colorectal cancer by facilitating nonhomologous end joining repair. Clin Cancer Res 25. 2019.

59. Xiang J, Fang L, Luo Y, Zhang J, Huang Y, Fan X, et al: Levels of human replication factor C4, a clamp loader, correlate with tumor progression and predict the prognosis for colorectal cancer. J Transl Med 12: 320, 2014.

60. Liu YZ, Wang BS, Jiang YY, Cao J, Hao JJ, Zhang Y, Xu X, Cai Y and Wang MR: MCMs expression in lung cancer: Implication of prognostic significance. J Cancer Sci 8: 3641-3647, 2017.

61. Behrens C, Solis LM, Lin H, Yuan P, Tang X, Kadara H, Riquelme E, Galindo H, Morán CA, Kalnhor N, et al: EZH2 protein expression associates with the early pathogenesis, tumor progression, and prognosis of non-small cell lung carcinoma. Clin Cancer Res 19: 6556-6565, 2013.

62. Shi R, Li M, Raghavan V, Tam S, Cabanero M, Pham NA, Shepherd FA, Moghal N and Tsao MS: Targeting the CDK4/6-Rb pathway enhances response to PI3K inhibition in PI3KCA-mutant lung squamous cell carcinoma. Clin Cancer Res 25: 5990-6000, 2018.

63. Chen L and Pan J: Dual cyclin-dependent kinase 4/6 inhibition by PD-032991 induces apoptosis and senescence in oesophageal squamous cell carcinoma cells. Br J Pharmacol 174: 2247-2434, 2017.

64. Ma Y, Lin D, Sun W, Xiao T, Yuan J, Han N, Guo S, Feng X, Su K, Mao Y, et al: Expression of targeting protein for xklp2 associated with both malignant transformation of respiratory epithelium and progression of squamous cell lung cancer. Clin Cancer Res 12: 1121-1126, 2006.

65. Mollinari C, Kleman JP, Jiang W, Schoehn G, Hunter T and Margolis RL: PRC1 is a microtubule binding and bundling protein essential to maintain the mitotic spindle midzone. J Cell Biol 157: 1175-1186, 2002.

66. Zhan P, Xi GM, Liu HB, Liu YF, Xu WJ, Zhu Q, Zhou ZJ, Miao YY, Wang XX, Jin JJ, et al: Protein regulator of cytokinesis-1 expression: Prognostic value in lung squamous cell carcinoma patients. J Thorac Dis 9: 2054-2060, 2017.

67. Huang J, Zhou Y, Thomas GS, Gu Z, Yang Y, Xu H, Tricot G and Zhan F: NEDD8 inhibition overcomes CKS1B-induced drug resistance by upregulation of p21 in multiple myeloma. Clin Cancer Res 21: 5532-5542, 2015.

68. Zhan F, Colla S, Wu X, Chen B, Stewart JP, Kuehl WM, Barlogie B and Shaughnessy JD Jr: CKS1B overexpressed in aggressive disease, regulates multiple myeloma growth and survival through SKP2- and p27kip1-dependent and -independent mechanisms. Blood 109: 4993-5001, 2007.

69. Romero Arenas MA, Whitsett TG, Aronova A, Henderson SA, LoBello J, Habra MA, Grubbs EG, Lee JE, Sircar K, Zamegarr R, et al: Protein expression of PTGT1 as a diagnostic biomarker in adenocortical carcinoma. Ann Surg Oncol 25: 801-807, 2018.

70. Guo XC, Li L, Gao ZH, Zhou HW, Li J and Wang QQ: The long non-coding RNA PTGT3P promotes growth and metastasis of cervical cancer through PTGT1. Aging (Albany NY) 11: 1333-1341, 2019.

71. Huang JL, Cao SW, Ou QS, Yang B, Zheng SH, Tang J, Chen J, Hu YW, Zheng L and Wang Q: The long non-coding RNA PTGT3P promotes cell growth and metastasis via up-regulating PTGT1 and activating PI3k/AKT signaling in hepatocellular carcinoma. Mol Cancer 17; 93, 2018.

72. You H, Lin H and Zhang Z: CKS2 in human cancers: Clinical roles and current perspectives (Review). Mol Clin Oncol 3: 459-463, 2015.

73. Hauge S, Nauke C, Hassvold G, Joel M, Rødland GE, Juvasens P, Stokke T and Syljuåsen RG: Combined inhibition of Weel and Chk1 gives synergistic DNA damage in S-phase due to distinct regulation of CDK activity and CDC45 loading. Oncotarget 8: 10966-10979, 2017.

74. Jensen JB, Munksgaard PP, Sørensen CM, Fristrup N, Birkenkamp-Demtroder K, Uhlof BS, Jensen KM, Orntoft TF and Dyrrskjøt L: High expression of karyopherin-α2 defines poor prognosis in non-muscle-invasive bladder cancer and in patients with invasive bladder cancer undergoing radical cystectomy. Eur Urol 59: 841-848, 2011.

75. Goto Y, Kurozumi A, Arai T, Nohata N, Kojima A, Kato M, Yamazaki K, Ishida Y, Naya Y, et al: Impact of novel miR-145-3p regulatory networks on survival in patients with castration-resistant prostate cancer. Br J Cancer 117: 409-420, 2017.

76. Garnett MJ, Mansfield J, Godwin C, Matsuoka T, Wu J, Russell P, Pines J and Venkitaraman AR: UBE2S elongates ubiquitin chains on APC/C substrates to promote mitotic exit. Nat Cell Biol 11: 1363-1369, 2009.

77. Akter KA, Hyodo T, Asano E, Sato N, Mansour MA, Ito S, Kato M, Yamazaki K, Ishida Y, Naya Y, et al: Impact of novel miR-145-3p regulatory networks on survival in patients with castration-resistant prostate cancer. Br J Cancer 117: 409-420, 2017.

78. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.