Identification of key genes and evaluation of clinical outcomes in lung squamous cell carcinoma using integrated bioinformatics analysis

YANGFENG SHI, YEPEING LI, CHAO YAN, HUA SU and KEJING YING

Department of Respiratory Medicine, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310000, P.R. China

Received May 12, 2019; Accepted September 2, 2019

DOI: 10.3892/ol.2019.10933

Abstract. Lung cancer is the leading cause of cancer-related mortality worldwide. Despite progress in the treatment of non-small-cell lung cancer, there are limited treatment options for lung squamous cell carcinoma (LUSC), compared with lung adenocarcinoma. The present study investigated the disease mechanism of LUSC in order to identify key candidate genes for diagnosis and therapy. A total of three gene expression profiles (GSE19188, GSE21933 and GSE74706) were analyzed using GEO2R to identify common differentially expressed genes (DEGs). The DEGs were then investigated using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis. A protein–protein interaction (PPI) network was constructed via the Search Tool for the Retrieval of Interacting Genes/Proteins, and visualized using Cytoscape software. The expression levels of the hub genes identified using CytoHubba were validated using the University of California, Santa Cruz (UCSC) database and the Human Protein Atlas. A Kaplan–Meier curve and Gene Expression Profiling Interactive Analysis were then employed to evaluate the associated prognosis and clinical pathological stage of the hub genes. Furthermore, non-coding RNA regulatory networks were constructed using the Gene-Cloud Biotechnology information website. A total of 359 common DEGs (155 upregulated and 204 downregulated) were identified, which were predominantly enriched in 'mitotic nuclear division', 'cell division', 'cell cycle' and 'p53 signaling pathway'. The PPI network consisted of 257 nodes and 2,772 edges, and the most significant module consisted of 66 upregulated genes. A total of 19 hub genes exhibited elevated RNA levels, and 10 hub genes had elevated protein levels compared with normal lung tissues. The upregulation of five hub genes (CCNB1, CEP55, FOXM1, MKI67 and TYMS; defined in Table I) were significantly associated with poor overall survival and unfavorable clinical pathological stages. Various ncRNAs, such as C1orf220, LINC01561 and MGC39584, may also play important roles in hub-gene regulation. In conclusion, the present study provides further understanding of the pathogenesis of LUSC, and reveals CCNB1, CEP55, FOXM1, MKI67 and TYMS as potential biomarkers or therapeutic targets.

Introduction

Lung cancer is one of the most prevalent cancer types and the leading cause of cancer-related mortality worldwide, accounting for an estimated 142,670 deaths in the USA in 2019 (1). Non-small cell lung cancer (NSCLC) constitutes 80‑85% of lung cancer cases, and the most common histological subtype is lung adenocarcinoma (LUAD), followed by lung squamous cell carcinoma (LUSC) (2). There has been much progress in the development of molecular LUAD-targeted therapies (3). Conversely, there has been limited progress in the development of treatments for LUSC, with the exception of immunotherapy (4‑6). It is hypothesized that LUAD originates from epithelial secretory cells, while LUSC originates from basal cells (2); furthermore, epidermal growth factor receptor, KRAS proto‑oncogene GTPase and EMAP‑like 4‑ALK receptor tyrosine kinase mutations tend to occur more frequently in LUAD (7). By contrast, LUSC exhibits more molecular abnormalities in the fibroblast growth factor receptor 1, phosphatidylinositol‑4,5‑bisphosphate 3‑kinase catalytic subunit alpha‑keratin 3 and discoidin domain receptor tyrosine kinase 2 genes (8). In summary, advances in the treatment, and the elucidation of the molecular mechanisms involved in LUAD far outweigh those made in LUSC. Due to the high incidence and mortality rate of LUSC, the need to reveal the pathogenesis, explore novel biomarkers and develop effective therapeutic strategies is imperative. Genomics, gene microarrays and high-throughput sequencing have been widely utilized in oncology research. Moreover, dysregulated gene expression plays a significant role in the pathogenesis and development of various malignancies.
role in cancer development. Several studies have identified hub genes from groups of differentially expressed genes (DEGs) based on integrated bioinformatics methods; using integrated bioinformatics analysis, Xia et al (9) identified anillin actin binding protein as a key gene in cervical cancer progression. Cui et al (10) also used bioinformatics analysis to demonstrate that maternally expressed 3 could function as a biomarker and predict the prognosis of breast cancer. Studies on NSCLC do exist (11,12), but most of these involve LUAD (13,14), and few are related to LUSC.

In the present study, three gene expression profiles for LUSC were downloaded from the Gene Expression Omnibus (GEO) database and GEO2R was utilized to screen DEGs between LUSC tissue samples and normal lung tissue samples. Subsequently, enrichment analysis, protein-protein interaction (PPI) network construction and module identification were performed to illustrate the significant associations between DEGs. Furthermore, hub genes from these DEGs were identified, validated and analyzed, revealing the prognostic and clinical values of the DEGS and ncRNA regulatory networks of hub genes involved in LUSC.

Materials and methods

Microarray data retrieval. The GEO (https://www.ncbi.nlm.nih.gov/geo/) database is a public, functional genomics repository of array and sequence-based, high-throughput gene expression data (15). In the present study, three gene expression profiles (GSE19188, GSE21933 and GSE74706) were downloaded from the GEO; GSE19188 contained 27 LUSC tissue samples and 65 normal lung tissue samples; GSE21933 contained 10 LUSC tissue samples and 10 matched normal lung tissue samples, and GSE74706 contained 8 LUSC tissue samples and 8 matched normal lung tissue samples. All probes were converted into their corresponding official gene symbols according to the annotation information provided each platform.

DEG screening. GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) is an online web tool that allows for the comparison between two groups of samples in different experimental conditions (16). It uses Bioconductor R packages to analyze selected datasets. In the present study, GEO2R was applied to screen for DEGs between LUSC tissue samples and normal lung tissue samples. The cut-off criteria were set as adjusted P-value <0.01 and llogFCi<2. DEGs common to the three datasets were selected for further analysis.

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs. The Database for Annotation, Visualization and Integrated Discovery (DAVID; https://david.ncifcrf.gov/) is a functional annotation tool that integrates biological data and analysis for multiple genes and proteins (17). GO and KEGG pathway enrichment analyses of DEGs were performed using DAVID. GO annotation includes biological process (BP), cellular component (CC) and molecular function (MF). P<0.05 was considered statistically significant.

PPI network construction and module identification. The PPI network of DEGs was constructed using the Search Tool for the Retrieval of Interacting Genes (STRING; https://string-db.org/cgi/input.pl?UserId=LUVvQCh21nYS&sessionId=XJ VX rbgj7wkb&input_page_show_search=on) database (18). A combined score ≥0.4 was defined as the cut-off point. Cytoscape software (version 3.7.1; https://cytoscape.org/) was then employed to visualize the PPI network (19). The most significant module was identified using Molecular Complex Detection (MCODE) (20), a plug-in of Cytoscape. The screening options were set as degree cut-off=2, node score cut-off=0.2, k-core=2 and Max. depth=100. GO and KEGG pathway enrichment analyses of genes in the most significant mode were subsequently performed using DAVID.

Hub gene selection and analysis. CytoHubba (21), a plugin of Cytoscape, was used to select hub genes of DEGs by identifying the intersection of the top 100 genes with 12 topological analysis methods. The biological process of hub genes was then analyzed and visualized by the Biological Networks Gene Oncology tool (BiNGO; http://apps.cytoscape.org/apps/bingo) plugin (22). The network of hub genes and their co-expression genes was constructed on the cBioportal (v3.0.6) platform (23).

Validation of hub genes. To validate the hub genes, RNA expression data from LUSC samples stored in The Cancer Genome Atlas database (TCGA) were visualized using USCS Xena (24). Next, the immunohistochemistry (IHC) results of the hub genes were verified on the Human Protein Atlas (HPA, https://www.proteinatlas.org/). The URLs that directly link to the images in the HPA are provided in Appendix S1. Hub genes with both higher RNA and protein expression levels in tumor tissue compared with normal lung tissue were selected for further analysis.

Survival analysis and clinical comparison of hub genes. The Kaplan-Meier (KM) plotter is an online tool that predicts the prognostic values of cancer-associated genes according to their expression levels (25). In the present study, analysis was restricted to squamous cell carcinomas, and patients were divided into two groups according to the expression levels of hub genes. Hazard ratio with 95% CI and log-rank P-value were calculated, and the hub genes with significant prognostic value were selected for further clinical comparison.

Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/) is an interactive web server for analyzing RNA sequencing expression data from TCGA and the Genotype-Tissue Expression (GTEx) project (26). The relevance between the expression of selected hub genes and the clinical TNM stage in LUSC was evaluated using data from the GEPIA database. P<0.05 was considered as the threshold.

Non-coding (nc)RNA regulatory network construction. Gene-Cloud Biotechnology information (GCBI; https://www.ncbi.com.cn/gclib/html/index) is a web tool that predicts the regulation of genes and ncRNAs, transcription factors and gene expression levels in disease. In the present study, GCBI was used to construct ncRNA regulatory networks of hub genes.

Results

Screening of DEGs in LUSC. A total of 3 gene expression profiles (GSE19188, GSE21933 and GSE74706) were
analyzed using GEO2R. Collectively, 359 common DEGs were identified (Fig. 1A), including 155 upregulated and 204 downregulated genes between normal lung tissue samples and tumor tissue samples in LUSC.

**GO and KEGG pathway enrichment analysis of DEGs.** To reveal the functions of the identified DEGs, GO and KEGG pathway enrichment analysis was performed using DAVID. The top 20 GO terms and the top 10 KEGG pathways are shown (Fig. 1B and C). GO annotation consisted of three groups: BP, CC and MF. The DEGs related to BP terms were predominantly enriched in ‘mitotic nuclear division’, ‘cell division’, ‘chromosome segregation’, ‘G1/S transition of mitotic cell cycle’ and ‘sister chromatid cohesion’. The DEGs related to CC terms were mainly enriched in ‘chromosome’, ‘centromeric region’, ‘condensed chromosome kinetochore’, ‘midbody’, ‘proteinaceous extracellular matrix’ and ‘collagen trimer’. The DEGs related to MF terms were mainly enriched in ‘protein binding’, ‘protein kinase binding’, ‘ATP binding’ and ‘signaling pattern recognition receptor activity’ and ‘scavenger receptor activity’ (Fig. 1B). The most enriched KEGG pathways were ‘cell cycle’, ‘p53 signaling pathway’, ‘ECM-receptor interaction’, ‘PPAR signaling pathway’ and ‘complement and coagulation cascades’ (Fig. 1C). These enriched terms indicate the pathogenic mechanisms of LUSC and provide direction for further research.

**PPI network construction, module identification and analysis.** To investigate the interactions between DEGs, the PPI dataset was downloaded from STRING and displayed using Cytoscape (Fig. 2). The PPI network consisted of 257 nodes and 2,772 edges. The most significant module was identified from the PPI network using the MCODE plugin, and comprised 66 nodes and 2,050 edges (Fig. 3A). The genes in the module were all upregulated in LUSC. Furthermore, GO and KEGG pathway enrichment analysis showed that these genes were primarily involved in ‘mitotic nuclear division’, ‘cell division’, ‘cell cycle’ and ‘p53 signaling pathway’ (Fig. 3B).

**Hub gene selection, validation and analysis.** A total of 19 hub genes were selected from the intersection between the top 100 genes, using 12 topological analysis methods in CytoHubba; the genes included AURKA, BIRC5, BUB1, CCNB1, CDK1, CDKN3, CEP55, EZH2, FOXM1, HJURP, HMMR, MKI67, NDC80, PBK, RFC4, TK1, TYMS and UBE2C (Table I). The biological process network of hub genes was constructed using BiNGO (Fig. 4A). The network of hub genes and their co-expression genes was constructed using cBioPortal (Fig. 4B). A heatmap of TCGA LUSC samples revealed that these hub genes could differentiate LUSC tissues from normal lung tissues (Fig. 4C), which was also consistent with the former result (Fig. 1A). For further validation, the protein levels of these hub genes appeared higher in LUSC tissue samples than in normal lung tissue samples, based on the data extracted from the HPA (Fig. 5); namely, AURKA, BIRC5, CCNB1, CDK1, CEP55, EZH2, FOXM1, MKI67, RFC4 and TYMS. The protein expression levels of these genes were also quantitatively analyzed using IHC (Fig. S1).

**Survival analysis and clinical comparison of hub genes.** The aforementioned 10 hub genes were further evaluated using the KM plotter to predict their prognostic values in LUSC (according to their expression levels). The results showed that higher mRNA expression levels of CCNB, CEP55, FOXM1, MKI67 and TYMS 1, in tumor vs. normal tissues, were all

Figure 1. Venn diagram showing GO and KEGG pathway enrichment analysis of DEGs in LUSC patients. (A) Overlapping areas show the common DEGs between datasets GSE19188, GSE21933 and GSE74706. (B) Top 20 GO terms from the GO enrichment analysis of DEGs. (C) Top 10 KEGG pathways in pathway enrichment analysis of DEGs. Rich factor=count/pop hits. GO, gene enrichment; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differently expressed genes; GO, gene ontology; LUSC, lung squamous cell carcinoma.
significantly associated with poor overall survival in LUSC patients (P<0.05; Fig. 6), whereas the others genes were not associated with prognosis. Multivariate analysis showed that CEP55, MKI67 and TYMS together with stage and sex had a significant impact on patient survival time (Table SI). In addition, the expression levels of the 5 hub genes were found to be associated with the clinical pathological stage, displayed as violin plots (P<0.05; Fig. 7). Higher expression levels of 5 hub genes tended to be associated with more advanced TNM stages, except for CCNB1, FOXM1 and MKI67 in stage IV. Perhaps the small sample size for stage IV disease could account for this finding. Moreover, ROC curves of the five hub genes (Fig. S2) exhibited a significant effect in distinguishing tumor tissues from normal tissues.

ncRNA regulatory network construction. The aforementioned enrichment analysis results revealed that the biological functions of the five selected hub genes were related to the cell cycle; these included ‘mitotic nuclear division’, ‘cell division’, ‘mitotic cytokinesis’, ‘regulation of cell cycle’, ‘G1/S transition of mitotic cell cycle’ and ‘G2/M transition of mitotic cell cycle’. Given the potent roles of ncRNAs in regulating biological processes, the related ncRNAs of the five hub genes were predicted using GCBI and exhibited as regulatory networks (Fig. 8). As shown, C1orf220, LINC01561 and MGC39584 simultaneously regulated CCNB1, CEP55, MKI67 and TYMS, indicating that these long non-coding RNAs (lncRNAs) play important roles in LUSC pathogenesis. The target regions of CCNB1 and MKI67 possess the same micro RNA (miRNA/miR) binding sites, which include miR-92a-3p, miR-559, miR-548a-5p and miR-548ab. CCNB1 and TYMS shared let-7b-5p, and MKI67 and CEP55 shared miR-16-5p, miR-192-5p and miR-215-5p. Moreover, the target regions of MKI67 and TYMS shared
Table I. Hub genes identified from the protein-protein interaction network using the Biological Networks Gene Oncology tool.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Official full name</th>
<th>Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AURKA</td>
<td>Aurora kinase A</td>
<td>Up</td>
</tr>
<tr>
<td>BIRC5</td>
<td>Baculoviral IAP repeat containing 5</td>
<td>Up</td>
</tr>
<tr>
<td>BUB1</td>
<td>BUB1 mitotic checkpoint serine/threonine kinase</td>
<td>Up</td>
</tr>
<tr>
<td>CCNB1</td>
<td>Cyclin B1</td>
<td>Up</td>
</tr>
<tr>
<td>CDK1</td>
<td>Cyclin dependent kinase 1</td>
<td>Up</td>
</tr>
<tr>
<td>CDKN3</td>
<td>Cyclin dependent kinase inhibitor 3</td>
<td>Up</td>
</tr>
<tr>
<td>CEP55</td>
<td>Centrosomal protein 55</td>
<td>Up</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of Zeste 2 polymeric repressive complex 2 subunit</td>
<td>Up</td>
</tr>
<tr>
<td>FOXM1</td>
<td>Forkhead Box M1</td>
<td>Up</td>
</tr>
<tr>
<td>HJURP</td>
<td>Holliday junction recognition protein</td>
<td>Up</td>
</tr>
<tr>
<td>HMMR</td>
<td>Hyaluronan mediated motility receptor</td>
<td>Up</td>
</tr>
<tr>
<td>MELK</td>
<td>Maternal embryonic Leucine zipper kinase</td>
<td>Up</td>
</tr>
<tr>
<td>MKI67</td>
<td>Marker of proliferation Ki-67</td>
<td>Up</td>
</tr>
<tr>
<td>NDC80</td>
<td>NDC80 kinetochore complex component</td>
<td>Up</td>
</tr>
<tr>
<td>PBK</td>
<td>PDZ binding kinase</td>
<td>Up</td>
</tr>
<tr>
<td>RFC4</td>
<td>Replication Factor C subunit 4</td>
<td>Up</td>
</tr>
<tr>
<td>TK1</td>
<td>Thymidine kinase 1</td>
<td>Up</td>
</tr>
<tr>
<td>TYMS</td>
<td>Thymidylate synthetase</td>
<td>Up</td>
</tr>
<tr>
<td>UBE2C</td>
<td>Ubiquitin conjugating enzyme E2 C</td>
<td>Up</td>
</tr>
</tbody>
</table>

Figure 3. GO and KEGG pathway enrichment analysis of the most significant module. (A) The most significant module obtained from the PPI network had 66 nodes and 2,050 edges. Red nodes represent upregulated genes. (B) Top 6 GO terms and KEGG pathways in enrichment analysis of the most significant module. GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PPI, protein-protein interaction.
Discussion

Although the incidence of lung cancer is declining, it is still responsible for the highest proportion of cancer-related mortality (1). Until now, there have been few specific therapeutics aimed at LUSC, compared with LUAD. Hence, it is imperative to identify novel biomarkers and effective therapeutic targets specific to LUSC.

In the present study, a total of 359 common DEGs were selected from three datasets, including 155 upregulated and 204 downregulated genes. The relative biological functions were primarily associated with ‘mitotic nuclear division’, ‘cell division’, ‘protein binding’ and ‘protein kinase binding’. KEGG pathways, including ‘cell cycle’, ‘p53 signaling pathway’, and ‘ECM-receptor interaction’, were dysregulated in LUSC. The PPI network determined the interactions of DEGs and a significant module was constructed and identified. GO and KEGG pathway enrichment analysis indicated that 66 genes from the most significant module in the PPI network were predominantly related to ‘mitotic nuclear division’, ‘cell division’, ‘cell cycle’ and ‘p53 signaling pathway’. Among the DEGs, 19 hub genes were selected. The co-expression network further validated the relationship between the hub genes, and revealed the pathways and potential therapeutic targets in which they are involved. In total, 10 of the 19 hub genes were validated as exhibiting elevated expressions levels of both mRNA and protein. Survival analysis revealed that high mRNA expression levels of CCNB1, CEP55, FOXM1, MKI67 and TYMS were related to poor overall survival. These five hub genes were also associated with advanced clinical pathological stages.
Cyclin B1 (CCNB1) is a pivotal member of the cyclin family that complexes with CDC2, exerting control over the cell cycle (27). As it gradually accumulates in the S phase, CCNB1 reaches its maximum level before mitosis and is then rapidly degraded in the M phase (27). Hence, CCNB1 is a key mediator of G2-M phase checkpoint surveillance.

Dysregulation of CCNB1 leads to cell hyperplasia and tumorigenesis, which has been reported in various cancers, including esophageal squamous cell carcinoma (28), breast cancer (29) and gastric cancer (30). Also, CCNB1 was highly expressed in NSCLC tissues compared with normal lung tissues (31,32). NSCLC patients with CCNB1 upregulation tend to have a poorer prognosis compared with patients with normal CCNB1 expression (32), particularly in LUSC (31). Furthermore, CCNB1 is also associated with long-term smokers and preneoplastic lesions (33), which could partially account for smoking being the leading cause of LUSC (34). Consistent with previous studies (27,35), the

![Image of Protein Levels](https://example.com/image.png)

Figure 5. Protein levels of the 10 hub genes were higher in tumor, compared with normal tissues. Immunohistochemistry of (A) AURKA (N: staining, not detected; intensity, negative; quantity, negative. T: staining, low; intensity, moderate; quantity, <25%) (B) BIRC5 (N: staining, not detected; intensity, weak; quantity, <25%) (C) CCNB1 (N: staining, not detected; intensity, negative; quantity, negative. T: staining, high; intensity, strong; quantity, 75-25%) (D) CDK1 (N: staining, not detected; intensity, negative; quantity, negative. T: staining, high; intensity, strong; quantity, 75-25%) (E) CEP55 (N: staining, not detected; intensity, negative; quantity, negative. T: staining, low; intensity, weak; quantity, 75-25%) (F) EZH2 (N: staining, not detected; intensity, negative; quantity, negative. T: staining, high; intensity, strong; quantity, >75%) (G) FOXM1 (N: staining, medium; intensity, weak; quantity, 75-25%) (H) MKI67 (N: staining, not detected; intensity, negative; quantity, negative. T: staining, high; intensity, strong; quantity, >75%) (I) RFC4 (N: staining, low; intensity, weak; quantity, >75%) (J) TYMS (N: staining, not detected; intensity, negative; quantity, negative. T: staining, medium; intensity, moderate; quantity, >75%) based on data from the Human Protein Atlas. N, normal tissue; T, tumor tissue. Gene definitions are displayed in Table I.
results of the present study revealed that CCNB1 was implicated in both the cell cycle and the p53-signaling pathway. Additionally, CCNB1 overexpression was associated with poor prognosis and advanced clinical pathological stages in LUSC in the present study. Yoshida et al (36) revealed that the upregulation of CCNB1 correlated with higher Ki-67 and PCNA in NSCLC, while CCNB1 and MKI67 (Ki-67) served as hub genes in the present study. Considering these findings, CCNB1 shows promise as a biomarker for LUSC diagnosis.
Centrosomal Protein 55 (CEP55), a highly coiled centrosomal protein, is an indispensable regulator of cytokinesis (37). During abscission, CEP55 recruits members of the endosomal-sorting complex to tear the cytokinetic bridge and divide the cytoplasm into two daughter cells (37,38). Cytokinetic disorders result in cellular transformation and

Figure 8. Non-coding RNA regulatory networks of the 5 hub genes. Related IncRNAs and targeted miRNAs regulatory networks of (A) CCNB1, (B) CEP55, (C) FOXM1, (D) MKI67 and (E) TYMS were constructed using Gene-Cloud Biotechnology information. IncRNAs, long non-coding RNAs; miRNAs, micro RNAs. Gene definitions are displayed in Table I.
malignancy (39). Certain studies have demonstrated that CEP55 upregulation contributes to different cancer types, such as breast cancer, ovarian cancer, colon cancer (39,40). CEP55 upregulation also promotes a number of events related to neoplasia, such as cell migration, invasion and anchorage independent growth (39,41). Kalimutho et al (39) revealed that CEP55 is implicated in the MEK1/2-MYC axis, which mediates aneuploidy and genomic instability in breast cancer. In NSCLC patients, elevated levels CEP55 promote migration and invasion via activation of the PI3K/AKT pathway (42), in addition to predicting unfavorable clinical outcomes (43). However, few studies have been conducted specifically on LUSC. GO enrichment analysis revealed that CEP55 was engaged in mitotic cytokinesis and mitotic nuclear division. CEP55 also exhibited a co-expression relationship with mitogen-activated protein kinase (MAPK) 1, which could implicate CEP55 in the MAPK-signaling pathway. In addition, CEP55 was highly expressed in LUSC and indicated poor clinical outcome. Accordingly, the findings of the present study indicate CEP55 as a potential therapeutic target.

Forkhead Box M1 (FOXM1) is a well-known transcription factor from the Forkhead family of proteins, which is upregulated in a broad range of tumors (44,45). As such, the FOXM1 regulatory network has been identified as a major predictor of unfavorable outcomes in several cancer types, such as breast cancer, colon cancer, prostate cancer (46). Similar to NSCLC, elevated expression levels of FOXM1 are significantly associated with poor prognosis (47). In LUSC, the findings of the present study resulted in similar conclusions. Expression of FOXM1 varies between different pathological stages; generally, FOXM1 mediates its pro-tumorigenic effect via transcriptional activation of its target (44). Accordingly, it has been revealed that FOXM1 regulates the expression levels of CCNB1, CEP55 and TYMS (48-50), which were identified as hub genes in the present study. Thus, it was demonstrated that FOXM1 plays a crucial role in LUSC tumorigenesis.

Marker of Proliferation Ki-67 (MKI67), a protein phosphatase 1-binding protein, is known as a cell proliferation marker in both laboratory and clinical cancer applications (51). MKI67 is degraded in the G1 and G2 phases, and gradually accumulates in the nucleoli following the onset of S phase (52). Moreover, MKI67 not only represents proliferation status, but also distinguishes rapid-growing from slow-growing tumors (51,53). In NSCLC, MKI67 has been defined as a diagnostic and prognostic marker (54). The results of the present study further validate that MKI67 upregulation indicates an advanced pathological stage and adverse overall survival time in LUSC. Consequently, the results of the present study support the crucial value of MKI67 in clinical diagnosis and during treatment.

Thymidylate Synthetase (TYMS) functions as a fundamental participant in thymidylate biosynthesis and de novo DNA replication (55). TYMS inhibition results in cell cycle arrest at the S phase (56), and high TYMS expression accelerates cell proliferation and leads to malignant behaviors in various types of solid tumor (56-58). In NSCLC, TYMS is upregulated in tumor tissues and linked to adverse prognosis (56). Further studies have revealed that TYMS is more highly expressed in LUSC than in LUAD (56,59). This is hypothesized to be responsible for the unsatisfactory treatment response to pemetrexed-based chemotherapy in LUSC (60). The present study revealed that TYMS was associated with the GO terms ‘G/S transition of mitotic cell cycle’ and ‘regulation of transcription’. Thus, the upregulation of TYMS may indicate poor prognosis and be pathologically detrimental. As indicated, TYMS disturbances contribute to LUSC development, and may therefore be a promising therapeutic target for further research.

With advances in next-generation deep sequencing, an increasing number of ncRNAs have been recognized in various diseases, including cancer (61). ncRNAs have the robust ability to regulate gene expression via versatile mechanisms that orchestrate biological processes. The present study assessed ncRNA regulatory networks to further investigate the effect of the hub genes in LUSC. Using, a competing endogenous RNA (ceRNA) network, Sui et al (62) reported that C1orf220 was upregulated in LUSC and indicated poor prognosis. Jiang et al (63) determined that the linc01561-miR-145-5p-MMP11 interaction contributed to the progression of breast cancer. However, the role of linc01561 in LUSC pathology is not fully understood, and MGC39584 (also known as LINC01667) remains to be investigated. Hub gene-related ncRNAs were revealed in the present study; C1orf220, LINC01561 and MGC39584 are all IncRNAs that simultaneously regulate the hub genes CCNB1, CEP55, MKI67 and TYMS and, hub genes sharing the same miRNAs may form a ceRNA network. Based on these findings, the present study suggests further research targets concerning the role of ncRNAs in LUSC.

RNA-based bioinformatics analyses could be considered a limitation of the present study. The verification of direct protein interactions, as well as functional experiments, would improve the validity of the results.

In conclusion, five hub genes (CCNB1, CEP55, FOXM1, MKI67 and TYMS) have been identified to play a central role in LUSC tumorigenesis, exhibiting ample diversity to the results derived from LUAD studies (13,14). Furthermore, ncRNAs such as C1orf220, LINC01561 and MGC39584 were shown to regulate these hub genes. This is the first proposal of key genes specifically upregulated in LUSC via integrated bioinformatics analysis. The findings of the present study may help inform the development of targeted therapeutics, though further experimental studies are required to verify these findings.

Acknowledgements

Not applicable.

Funding

This study was supported by a grant from the National Natural Scientific Foundation of China (grant no. 31627801).

Availability of data and materials

ONCOLOGY LETTERS

Authors' contributions
YS conducted the study and wrote the manuscript.YL applied statistical, computational and other techniques to synthesize and interpret the data, and revised the manuscript. CY and HS downloaded and analyzed the data. KY designed and directed the study.

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