Comprehensive Analysis of The Expression and Prognosis For KRTs in LUSC Based on Bioinformatics Analysis

Qiangqiang Zheng  
Sichuan University

Kaili Huang  
Sichuan University

Shihui Min (minshihui@wchfscu.cn)  
Sichuan University

Research Article

Keywords: Bioinformatics analysis, keratins, prognosis, squamous cell carcinoma, survival

DOI: https://doi.org/10.21203/rs.3.rs-775034/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Keratins (KRTs), a family of genes that encode a series of intermediate filament proteins, which expressed in epithelial cells of various tissues and had been identified as being involved in various tumors. Previous evidences indicated that KRTs were implicated in tumorigenesis and development of lung cancer. However, a comprehensive analysis of KRTs in lung squamous cell carcinoma (LUSC) and their roles in the tumorigenesis and progression of this disease is lacking. Therefore, we investigated the transcriptional level, proteomic level, and prognosis of KRTs in patients with LUSC from TCGA, Gene Expression Profiling Interactive Analysis (GEPIA), Human Protein Atlas (HPA), Kaplan-Meier Plotter, and cBioPortal databases. We confirmed that the expression levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 were higher in LUSC tissues than in adjacent normal lung tissues. Survival analyses using Kaplan-Meier Plotter database and TCGA database revealed that the high expression levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 were associated with poor prognosis in patients with LUSC. ROC curves indicated that KRTs had poor roles in predicting the overall survival (OS) of LUSC patients. KRTs and their 50 most frequently altered neighbor genes were mainly enriched in epidermis development, keratinization, keratinocyte differentiation, Rap1 signaling pathway, and transcriptional misregulation in cancer. The results of clinical data analyses indicated that the expression levels of KRT13, KRT16, KRT17, and KRT19 were positively correlated with nodal metastasis, the expression level of KRT15 and KRT17 were positively correlated with primary tumor, and the high expression level of KRT15 was positively correlated with TNM stage. Our study may provide new theoretical basis and research direction for the discovery of potential therapeutic targets and new prognostic biomarkers in patients with LUSC.

Background

Lung cancer is a familiar clinical malignant tumor, which ranks the top among tumors worldwide in terms of morbidity and mortality. In January 2021, the International Cancer Research Institute (IARC) published the latest cancer statistics of 36 cancers in 185 countries, and revealed that the new cases of lung cancer reached 2.2 million, ranking second in all new cancer population, and 1.8 million people died of lung cancer, ranking the first in cancer deaths. In China, the new cases and mortality of lung cancer all ranked first (820 thousand and 710 thousand respectively), accounting for 17.9% and 23.8% of new cases and mortality of all cancers.

- Non-small cell lung cancer (NSCLC) approximately accounts for 85% of all lung cancer, and is associated with a poor 5-year overall survival (OS) rate of 19.8%. The commonest subtypes of lung cancer are lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). LUSC is most strongly associated with smoking in a dose-dependent manner, and is more common in men. Generally LUSC grows more fast with larger masses than LUAD of the same stage, and tends to nodal metastasis at the early stages. Pan-cancer studies have shown that the molecular mechanisms of carcinogenesis and progression between LUAD and LUSC are highly heterogeneous, therefore, therapeutic strategies for LUAD are often invalid for LUSC. In recent years, with the continuous development of molecular biology and advocacy of precision medicine, the development of treatment targets and targeted drugs in lung cancer are becoming more and more mature, and the survival time of patients with lung cancer is slightly prolonged. Previous studies have reported that numerous genes were related to LUAD, including EGFR, AKT1, KRAS, TP53, PTEN, and others. However, the research on LUSC lacks valuable gene mutation and the progress is slow, there is no valuable targeted drug for the time being.

LUSC is a product of cumulative genetic and proteomic aberrations. The keratin family was believed to be closely associated with the carcinogenesis and progression of LUSC. Keratins (KRTs) are essential components of intermediate fibrin in the cytoplasm of epithelial cells. The expression levels of KRTs in cytoplasm reflects the differentiation of cells, and the abnormal expression of KRTs has a significant effect on the biological behavior of tumor cells. Previous studies have discovered that the expression of KRTs was deregulated in mutiple human malignancies, including lung cancer, breast...
cancer\textsuperscript{17}, and colon cancer\textsuperscript{18}. In addition, the expression levels of KRTs are distinct in different subtypes of lung cancer. For example, compared with normal lung tissues, the expression level of KRT5 was significantly upregulated in LUSC tissues, but not in LUAD tissues and small cell lung cancer (SCLC) tissues\textsuperscript{19,20}.

The abnormal expressions of KRTs and their relationship with clinicopathological features and prognosis in patients with LUSC have been partly reported. However, the specific roles of KRTs in LUSC are not completely clear, and there is no systematic evaluation to study the expression levels of KRTs in LUSC, or to assess whether their expression levels are related to the clinical characteristics and progression of this cancer. Previous studies have generated a large volume of differentially expressed genes (DEGs) and microarray data, which makes it possible to search for gene function prediction\textsuperscript{21}, disease-associated biomarkers\textsuperscript{22}, and alternative splicing\textsuperscript{23} by analyzing these data. In present study, we analyzed the expression levels of KRTs in patients with LUSC in detail to determine the expression pattern, potential function, and diagnostic and prognostic value of KRTs in LUSC.

Materials And Methods

Data collection and processing. High-throughput gene expression data of LUSC tissues and paracancerous lung tissues were downloaded from the Cancer Genome Atlas (TCGA) data portal (https://tcga-data.nci.nih.gov/tcga). These RNA-seq data (HTSeq-count) from Illumina HiSeq RNASeq platform consisted of 502 LUSC samples and 49 adjacent normal lung samples, and were achieved from the publicly available Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/).

Exploration of the aberrantly expressed KRTs in LUSC. In order to identify the aberrantly expressed KRTs, the ‘limma’ package in RStudio (version 1.4.1106) was applied to compare the expression profiles of LUSC tissues with those of adjacent lung tissues\textsuperscript{24}. The $|\log_2\text{fold change [FC]}| > 2$, \(P\) value < 0.01, and false discovery rate (FDR) < 0.01 were considered as the threshold values for DEGs identification. We selected the KRTs for our study, and used ‘gplots’ and ‘ggplot2’ packages of RStudio software to produce the heatmap and volcanoplot of the differentially expressed KRTs.

The GEPIA database. To determine the difference of the transcriptional level of KRTs between LUSC tissues and normal lung tissues, we used the "Expression analysis-Profile" and "Expression analysis-Box Plots" modules of the Gene Expression Profiling Interactive Analysis 2 (GEPIA2) web server (http://gepia2.cancer-pku.cn/#analysis) to obtain scatter diagrams and box plots of the transcriptional level between the LUSC tissues and the adjacent normal lung tissues, under the settings of P-value cutoff = 0.01, $|\log_2\text{FC}| = 1$, and “Match TCGA normal and GTEx data”\textsuperscript{25}.

The HPA database. The Human Protein Atlas (HPA) database (https://www.proteinatlas.org/) is a database of immunohistochemistry (IHC) - based protein expression profiles in cancer tissues, normal tissues, and cell lines\textsuperscript{26}. The KRTs protein expression IHC images in clinical specimens of LUSC patients were obtained from this database.

Survival analysis. Firstly, the Kaplan-Meier plotter database (http://kmplot.com) was used to evaluate the survival effects of KRTs, which was able to assess the impact of 54,000 genes on the prognosis of 21 cancers, including breast cancer, ovarian cancer, lung cancer, and gastric cancer\textsuperscript{27}. In addition, LUSC patients obtained from TCGA database were divided into the high expression group and the low expression group based on the median expression of KRTs, and the survival curves were produced with ‘hash’ and ‘survival’ packages of RStudio.

Diagnostic role of the KRTs in LUSC. The receiver operating characteristic (ROC) curve was proformed to assess the diagnostic roles of KRTs in patients with LUSC. ROC curve analysis was performed in RStudio using procedures from the ‘pROC’ package\textsuperscript{28}.

Potential molecular mechanism of KRTs in LUSC. It is speculated that the abnormal expression of KRTs in LUSC may be closely related to genetic alterations, including amplification, deletion, or point mutations. Therefore, cBioPortal was used to summarize the possible genetic changes of KRTs in LUSC\textsuperscript{29}.


**Co-expression and network of KRTs in LUSC.** Gene expression data of LUSC tissues and adjacent normal lung tissues were extracted from the TCGA data portal. We selected the gene expression data of KRTs to study. The co-expression analysis of KRTs was performed in RStudio using procedures from the ‘pheatmap’ package. Relative coefficient > 0.3 was considered statistically significant.

**PPI network and functional enrichment analysis of KRTs and altered neighbor genes.** Based on the GEPIA database (http://gepia2.cancer-pku.cn/#analysis), we constructed a network of KRTs and 50 most frequently changed neighboring genes. The selected genes were input into the online search tool for the retrieval of interacting genes (STRING, version 11.0) database to evaluate the interaction among the genes\(^{30}\). Interactions with a combined score > 0.4 were defined as statistically significant. Cytoscape software (version 3.8.2) was used to visualize the integrated regulatory networks and to mark with different colors according to the degree calculated by CytoHubba (version 0.1) plugin\(^ {31}\). To identify the pathways which had the most significant involvement with the KRTs, all genes were input into the Database for Annotation Visualization and Integrated Discovery (DAVID)\(^ {32}\) for gene ontology (GO) analysis, which included the biological process (BP), cellular component (CC), and molecular function (MF)\(^ {33}\), and Kyoto Encyclopedia of Genes and Genomes (KEGG)\(^ {30}\) pathway analysis enrichment, and then visualized in bubble chart by ‘ggplot2’ package of RStudio. \(P < 0.05\) and gene count \(\geq 5\) were considered to indicate a significant difference in the functional enrichment analysis. In addition, the molecular complex detection (MCODE, version 2.0.0) plugin of Cytoscape software was used to identify the important molecules in PPI networks\(^ {34}\). The GO analysis of KRTs was analyzed and plotted using ClueGO (version 2.5.7) and CluePedia (version 1.5.7)\(^ {35}\).

**Relationship between KRTs and clinicopathological features in LUSC.** The clinical data of LUSC was extracted from the TCGA data portal. Gene expression data and clinical data were combined by ‘hash’ package of RStudio, and the clinical data was divided into high expression group and low expression group according to the median expression. The relationships between the expression levels of KRTs and clinical characteristics were analyzed by statistical analysis of high expression group and low expression group, and were visualized in forestplot by ‘forestplot’ package of RStudio. Statistical analyses were performed using Pearson's chi-squared test in SPSS (version 22.0). The odds ratios (ORs) and corresponding 95% confidence intervals (CIs) were calculated for each clinical parameter by STATA14. The cutoffs of \(P\) value was defined as 0.05.

**Results**

**Aberrantly expressed KRTs based on TCGA database in LUSC.** The expression levels of each gene transformed with log2 were calculated by DESeq. Following the calculating criteria, we achieved 6,703 aberrantly expressed genes between LUSC tissues and normal lung tissues, including 4,983 highly and 1,720 lowly expressed genes. Then we listed the 57 KRTs according to the value of \(|\log_2 FC|\) (Table 1), which demonstrated that KRTs were upregulated in LUSC. The ‘gplots’ and ‘ggplot2’ packages of RStudio were used to produce heatmap and volcanoplot of the 57 KRTs (Fig. 1).
Table 1
The aberrantly expressed KRTs in LUSC based on TCGA database

| Gene symbol | LogFC  | P-value   | FDR       |
|------------|--------|-----------|-----------|
| KRT1       | 3.150  | 4.04E-09  | 1.07E-08  |
| KRT10      | 2.627  | 4.42E-19  | 2.32E-18  |
| KRT13      | 7.657  | 1.82E-32  | 1.93E-31  |
| KRT14      | 8.642  | 1.65E-33  | 1.83E-32  |
| KRT15      | 6.278  | 2.61E-45  | 4.72E-44  |
| KRT16      | 8.019  | 7.89E-48  | 1.57E-46  |
| KRT16P4    | 5.377  | 4.46E-21  | 2.64E-20  |
| KRT16P5    | 5.269  | 2.86E-18  | 1.43E-17  |
| KRT16P6    | 6.226  | 1.33E-20  | 7.66E-20  |
| KRT17      | 6.686  | 2.07E-58  | 5.65E-57  |
| KRT17P1    | 5.444  | 1.31E-40  | 1.99E-39  |
| KRT17P2    | 4.824  | 1.20E-35  | 1.48E-34  |
| KRT17P3    | 6.956  | 3.81E-31  | 3.79E-30  |
| KRT17P6    | 5.916  | 5.57E-45  | 9.93E-44  |
| KRT19      | 2.283  | 9.96E-26  | 7.58E-25  |
| KRT19P1    | 2.816  | 7.94E-23  | 5.21E-22  |
| KRT2       | 3.249  | 4.70E-10  | 1.34E-09  |
| KRT20      | 5.796  | 2.16E-13  | 7.91E-13  |
| KRT23      | 3.906  | 3.80E-16  | 1.67E-15  |
| KRT24      | 6.596  | 3.38E-13  | 1.22E-12  |
| KRT25      | 4.112  | 1.55E-07  | 3.56E-07  |
| KRT28      | 2.792  | 0.001     | 0.002270644 |
| KRT3       | 5.407  | 4.07E-29  | 3.69E-28  |
| KRT31      | 9.879  | 1.56E-34  | 1.82E-33  |
| KRT32      | 4.445  | 7.49E-26  | 5.74E-25  |
| KRT33A     | 8.790  | 5.43E-29  | 4.89E-28  |
| KRT33B     | 5.453  | 1.58E-18  | 8.07E-18  |
| KRT34      | 6.314  | 7.04E-34  | 7.96E-33  |
| KRT35      | 4.117  | 6.51E-10  | 1.84E-09  |
| KRT36      | 4.168  | 1.56E-14  | 6.17E-14  |

*Note: P < 0.05 was considered significant.*

Abbreviations: FDR, false discovery rate; KRTs, keratins; LUSC, lung squamous cell carcinoma; TCGA, the cancer genome atlas.
| Gene symbol | LogFC | P value | FDR     |
|------------|-------|---------|---------|
| KRT37      | 5.510 | 3.10E-20 | 1.75E-19|
| KRT38      | 5.168 | 1.38E-13 | 5.12E-13|
| KRT40      | 3.182 | 1.89E-07 | 4.32E-07|
| KRT41P     | 5.153 | 4.56E-18 | 2.26E-17|
| KRT42P     | 4.480 | 1.67E-22 | 1.07E-21|
| KRT43P     | 4.956 | 3.46E-12 | 1.16E-11|
| KRT5       | 7.382 | 4.95E-65 | 1.64E-63|
| KRT6A      | 8.174 | 5.28E-67 | 1.85E-65|
| KRT6B      | 8.109 | 2.00E-47 | 3.92E-46|
| KRT6C      | 8.138 | 1.53E-35 | 1.87E-34|
| KRT71      | 2.183 | 3.68E-10 | 1.06E-09|
| KRT74      | 6.886 | 9.26E-52 | 2.09E-50|
| KRT75      | 7.961 | 3.93E-30 | 3.74E-29|
| KRT76      | 5.765 | 1.67E-15 | 7.03E-15|
| KRT77      | 8.284 | 6.97E-29 | 6.22E-28|
| KRT78      | 5.904 | 1.19E-33 | 1.33E-32|
| KRT80      | 2.049 | 8.44E-16 | 3.62E-15|
| KRT81      | 5.901 | 8.82E-19 | 4.56E-18|
| KRT82      | 3.436 | 8.75E-15 | 3.51E-14|
| KRT83      | 4.985 | 3.68E-18 | 1.83E-17|
| KRT84      | 6.829 | 7.24E-22 | 4.49E-21|
| KRT85      | 3.122 | 2.25E-08 | 5.58E-08|
| KRT86      | 2.280 | 2.40E-09 | 6.48E-09|
| KRT8P30    | 2.394 | 2.14E-11 | 6.79E-11|
| KRT8P35    | 2.663 | 1.70E-19 | 9.17E-19|
| KRT8P38    | 2.898 | 8.09E-09 | 2.09E-08|
| KRT9       | 7.137 | 4.56E-37 | 5.94E-36|

Note: P < 0.05 was considered significant.

Abbreviations: FDR, false discovery rate; KRTs, keratins; LUSC, lung squamous cell carcinoma; TCGA, the cancer genome atlas.
Transcriptional level of KRTs in patients with LUSC. Using the GEPIA database, we compared the transcriptional levels of KRTs between LUSC tissues and normal lung tissues. The results indicated that the transcriptional levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 were higher in LUSC tissues than in normal lung tissues (Fig. 2).

Protein expression level of KRTs in patients with LUSC. To determine the difference of KRTs protein expression between LUSC tissues with normal lung tissues, we obtained IHC staining images of KRTs from HPA database (Fig. 3). Consistent with the above results of KRTs mRNA expression, the results showed that the protein levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 were higher in LUSC tissues than in normal lung tissues.

Association of the KRTs with the prognosis of patients with LUSC. The prognosis of LUSC patients was analyzed using the Kaplan-Meier survival plotter database. The results indicated that low expression levels of KRT5 [HR = 1.27 (1.12–1.44), P = 0.00018], KRT6A [HR = 1.39 (1.23–1.58), P = 2.7e-07], KRT6B [HR = 1.39 (1.22–1.58), P = 3.4e-07], KRT6C [HR = 1.43 (1.26–1.63), P = 2.3e-08], KRT13 [HR = 1.17 (1.03–1.32), P = 0.016], KRT14 [HR = 1.16 (1.02–1.32), P = 0.022], KRT15 [HR = 1.15 (1.01–1.30), P = 0.036], KRT16 [HR = 1.55 (1.36–1.76), P = 1.5e-11], KRT17 [HR = 1.44 (1.27–1.64), P = 1.3e-08], KRT19 [HR = 1.30 (1.15–1.48), P = 4e-05], and KRT23 [HR = 1.20 (1.06–1.37), P = 0.0039] were correlated with significantly better overall survival (OS) in LUSC patients (Fig. 4A), low expression levels of KRT5 [HR = 1.25 (1.03–1.51), P = 0.024], KRT6A [HR = 1.37 (1.13–1.66), P = 0.0012], KRT6B [HR = 1.42 (1.18–1.72), P = 0.00029], KRT6C [HR = 1.58 (1.30–1.91), P = 2.8e-06], KRT13 [HR = 1.34 (1.11–1.62), P = 0.0027], KRT14 [HR = 1.85 (1.52–2.25), P = 3.6e-10], KRT15 [HR = 1.96 (1.61–2.39), P = 7.7e-12], KRT16 [HR = 1.34 (1.10–1.62), P = 0.0029], KRT17 [HR = 1.35 (1.12–1.64), P = 0.0019], KRT19 [HR = 1.44 (1.18–1.74), P = 0.00022], and KRT23 [HR = 1.33 (1.10–1.61), P = 0.0036] were correlated with significantly better progression-free survival (PFS) in LUSC patients (Fig. 4B).

In addition, we downloaded the survival data of patients with LUSC from TCGA database, combined the survival data and gene expression data, and divided them into high expression group and low expression group to study the effect of KRTs on the prognosis of patients with LUSC, the results showed that low expression levels of KRT5 (P = 0.00493), KRT6A (P = 0.00163), KRT6B (P = 2e-05), KRT6C (P = 0.00026), KRT13 (P = 0.01032), KRT14 (P = 0.00121), KRT15 (P = 0.00074), KRT16 (P = 0.02083), KRT17 (P = 9e-04), KRT19 (P = 0.00027), and KRT23 (P = 0.00111) were predicted to have significantly better OS in LUSC patients (Fig. 4C), which were consistent with the results of Kaplan-Meier survival plotter database.

ROC of the KRTs in patients with LUSC. We obtained the gene expression data and clinical data from TCGA database. ROC curve analysis was performed in RStudio using procedures from the ‘PROC’ package. The results indicated that the KRTs had different specificity and sensitivity in predicting the prognosis of LUSC patients (Fig. 5). The area under receiver operating characteristic (AUC) of KRT5 was 60.753, which was the biggest of all KRTs. The sensitivity and specificity of KRT13 and KRT23 were 74.194 and 86.31, respectively, which were the biggest of all KRTs.

Potential molecular mechanism of the KRTs in patients with LUSC. We used the cBioPortal online tool for TCGA LUSC cohort to analyze the alterations of KRTs. The mutation, which can change the function of protein by changing gene sequence, is the most common gene alteration in KRTs (Fig. 6A). In the current study, the results showed that 13% (59/469) cases had genetic alterations, including inframe mutation, missense mutation, splice mutation, truncating mutation, amplification, and deep deletion (Fig. 6B). The KRT23 (2.8%) was the most common mutational gene among the KRTs, including missense mutation, splice mutation, amplification, and deep deletion.

The co-expression of KRTs in patients with LUSC. TCGA database was used to analyze the expression data of KRTs, and RStudio software was used to calculate the correlation between them, and Pearson’s correction was performed. The heatmap was shown in Fig. 7. The results indicated that there were significant and positive correlations between the following KRTs: KRT5 with KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, and KRT19; KRT6A with KRT6B, KRT6C, KRT14, KRT16, KRT17, and KRT19; KRT6B with KRT6C, KRT14, KRT16, and KRT17; KRT6C with KRT14, KRT16, and KRT17; KRT13 with KRT15 and KRT19; KRT14 with KRT16 and KRT17; KRT15 with KRT19; and KRT16 with KRT17.
GO function and KEGG pathway analyses for KRTs and their related genes. We constructed the network for KRTs and their 50 most frequently altered neighbor genes, and found that there were 262 nodes and 1,121 edges in this network (Fig. 8A). The results of functions of 262 genes were shown in Table 2. GO enrichment analysis predicted the functional roles of 262 genes on the basis of three aspects, including biological process (BP) terms, cell component (BP) terms, and molecular function (BP) terms. We found that the 262 genes were mainly enriched in epidermis development, keratinization, and keratinocyte differentiation in BP terms (Fig. 8B), enriched in cornified envelope, extracellular exosome, and desmosome in CC terms (Fig. 8B), and enriched in structural molecule activity, structural constituent of cytoskeleton, and cadherin binding involved in cell-cell adhesion in MF terms (Fig. 8B). KEGG pathway analysis demonstrated that the 262 genes were enriched in P53 signaling pathway, drug metabolism-cytochrome P450, Rap1 signaling pathway, and transcriptional misregulation in cancer (Fig. 8C).
Table 2
GO analysis and KEGG pathway analysis of the KRTs and their neighbor genes associated with LUSC

| Category | GO ID     | Term                                              | Count | P value     |
|----------|-----------|---------------------------------------------------|-------|-------------|
| BP       | GO:0008544| epidermis development                             | 23    | 1.56E-22    |
| BP       | GO:0031424| keratinization                                     | 14    | 4.53E-14    |
| BP       | GO:0030216| keratinocyte differentiation                       | 16    | 7.93E-14    |
| BP       | GO:0018149| peptide cross-linking                              | 12    | 5.08E-11    |
| BP       | GO:0016337| single organismal cell-cell adhesion               | 10    | 9.14E-06    |
| BP       | GO:0031581| hemidesmosome assembly                             | 5     | 1.48E-05    |
| BP       | GO:0007155| cell adhesion                                      | 18    | 1.76E-04    |
| BP       | GO:0098609| cell-cell adhesion                                 | 13    | 3.30E-04    |
| BP       | GO:0042060| wound healing                                      | 7     | 7.46E-04    |
| BP       | GO:0043588| skin development                                   | 5     | 0.001       |
| BP       | GO:0010951| negative regulation of endopeptidase activity     | 8     | 0.001       |
| BP       | GO:0007156| homophilic cell adhesion via plasma membrane adhesion molecules | 9 | 0.001 |
| BP       | GO:0007010| cytoskeleton organization                          | 8     | 0.006       |
| BP       | GO:0050680| negative regulation of epithelial cell proliferation | 5   | 0.007       |
| BP       | GO:007605 | sensory perception of sound                        | 7     | 0.009       |
| BP       | GO:006508 | proteolysis                                       | 14    | 0.019       |
| BP       | GO:0060070| canonical Wnt signaling pathway                    | 5     | 0.026       |
| BP       | GO:0030036| actin cytoskeleton organization                    | 6     | 0.032       |
| BP       | GO:0016055| Wnt signaling pathway                              | 7     | 0.042       |
| BP       | GO:0006810| transport                                         | 10    | 0.047       |
| CC       | GO:001533 | cornified envelope                                 | 14    | 1.99E-14    |
| CC       | GO:0070062| extracellular exosome                              | 86    | 2.49E-14    |
| CC       | GO:0030057| desmosome                                         | 10    | 1.22E-11    |
| CC       | GO:005882 | intermediate filament                              | 13    | 3.46E-08    |
| CC       | GO:005913 | cell-cell adherens junction                        | 18    | 1.60E-06    |
| CC       | GO:005911 | cell-cell junction                                 | 13    | 3.32E-06    |
| CC       | GO:005886 | plasma membrane                                    | 85    | 1.08E-05    |
| CC       | GO:0016324| apical plasma membrane                             | 13    | 5.38E-04    |
| CC       | GO:005576 | extracellular region                               | 38    | 7.04E-04    |
| CC       | GO:005737 | cytoplasm                                         | 92    | 0.002       |

Note: P < 0.05 was considered significant.

Abbreviations: BP, biological process; CC, cellular component; GO, gene ontology; KEGG, kyoto encyclopedia of genes and genomes; KRTs, keratins; LUSC, lung squamous cell carcinoma; MF, molecular function.
| Category | GO ID       | Term                                           | Count | P value  |
|----------|-------------|------------------------------------------------|-------|----------|
| CC       | GO:0005887  | integral component of plasma membrane          | 33    | 0.002    |
| CC       | GO:0005856  | cytoskeleton                                    | 13    | 0.004    |
| CC       | GO:0043005  | neuron projection                               | 10    | 0.005    |
| CC       | GO:0009986  | cell surface                                    | 16    | 0.006    |
| CC       | GO:0005615  | extracellular space                             | 30    | 0.007    |
| CC       | GO:0045095  | keratin filament                                | 6     | 0.011    |
| CC       | GO:0031225  | anchored component of membrane                  | 6     | 0.017    |
| CC       | GO:0030018  | Z disc                                         | 6     | 0.021    |
| CC       | GO:0015629  | actin cytoskeleton                              | 8     | 0.027    |
| CC       | GO:001726   | ruffle                                         | 5     | 0.032    |
| CC       | GO:0016323  | basolateral plasma membrane                     | 7     | 0.033    |
| CC       | GO:0005925  | focal adhesion                                  | 11    | 0.036    |
| CC       | GO:0030054  | cell junction                                   | 12    | 0.043    |
| CC       | GO:0042470  | melanosome                                      | 5     | 0.046    |
| MF       | GO:0005198  | structural molecule activity                    | 26    | 6.41E-15 |
| MF       | GO:0005200  | structural constituent of cytoskeleton          | 12    | 3.37E-07 |
| MF       | GO:0098641  | cadherin binding involved in cell-cell adhesion | 15    | 5.06E-05 |
| MF       | GO:0004252  | serine-type endopeptidase activity             | 12    | 8.28E-04 |
| MF       | GO:0005509  | calcium ion binding                             | 20    | 0.005    |
| MF       | GO:0004867  | serine-type endopeptidase inhibitor activity    | 6     | 0.011    |
| MF       | GO:0008236  | serine-type peptidase activity                 | 5     | 0.011    |
| MF       | GO:0003924  | GTPase activity                                 | 9     | 0.016    |
| MF       | GO:0030674  | protein binding, bridging                       | 5     | 0.022    |
| KEGG     | hsa04115    | p53 signaling pathway                           | 6     | 0.001    |
| KEGG     | hsa00982    | Drug metabolism-cytochrome P450                 | 5     | 0.010    |
| KEGG     | hsa04015    | Rap1 signaling pathway                          | 8     | 0.016    |
| KEGG     | hsa05202    | Transcriptional misregulation in cancer         | 7     | 0.018    |

Note: P < 0.05 was considered significant.

Abbreviations: BP, biological process; CC, cellular component; GO, gene ontology; KEGG, kyoto encyclopedia of genes and genomes; KRTs, keratins; LUSC, lung squamous cell carcinoma; MF, molecular function.
| Category | GO ID   | Term                                | Count | P value |
|----------|---------|-------------------------------------|-------|---------|
| KEGG     | hsa04974| Protein digestion and absorption    | 5     | 0.024   |
| KEGG     | hsa04540| Gap junction                         | 5     | 0.024   |
| KEGG     | hsa05200| Pathways in cancer                   | 11    | 0.024   |
| KEGG     | hsa05146| Amoebiasis                           | 5     | 0.043   |

Note: *P* < 0.05 was considered significant.

Abbreviations: BP, biological process; CC, cellular component; GO, gene ontology; KEGG, kyoto encyclopedia of genes and genomes; KRTs, keratins; LUSC, lung squamous cell carcinoma; MF, molecular function.

In addition, we use the MCODE plugin in Cytoscape to construct the module with the highest MCODE score, which contains 41 nodes and 320 edges (Fig. 8D). Subsequent GO function analysis revealed that the 41 genes were mainly enriched in epidermis development, keratinization, keratinocyte differentiation, intermediate filament, desmosome, cornified envelope, structural molecule activity, structural constituent of cytoskeleton, and calcium ion binding (Table 3, Fig. 8E).
| Category | GO ID       | Term                                              | Count | P value   |
|----------|-------------|---------------------------------------------------|-------|-----------|
| BP       | GO:0008544  | epidermis development                            | 15    | 3.19E-23  |
| BP       | GO:0031424  | keratinization                                    | 10    | 1.14E-15  |
| BP       | GO:0030216  | keratinocyte differentiation                      | 10    | 9.28E-14  |
| BP       | GO:0018149  | peptide cross-linking                             | 8     | 1.92E-11  |
| BP       | GO:0031581  | hemidesmosome assembly                            | 5     | 1.21E-08  |
| BP       | GO:0016337  | single organismal cell-cell adhesion              | 7     | 1.13E-07  |
| BP       | GO:0007155  | cell adhesion                                     | 8     | 7.85E-05  |
| BP       | GO:0007156  | homophilic cell adhesion via plasma membrane adhesion molecules | 5     | 4.80E-04  |
| BP       | GO:0007010  | cytoskeleton organization                         | 5     | 5.16E-04  |
| CC       | GO:0005882  | intermediate filament                             | 12    | 4.53E-16  |
| CC       | GO:0030057  | desmosome                                         | 8     | 3.92E-14  |
| CC       | GO:0001533  | cornified envelope                                | 9     | 5.03E-14  |
| CC       | GO:0070062  | extracellular exosome                             | 26    | 7.55E-12  |
| CC       | GO:0045095  | keratin filament                                  | 6     | 2.23E-06  |
| CC       | GO:0005911  | cell-cell junction                                | 5     | 4.87E-04  |
| CC       | GO:0005913  | cell-cell adherens junction                       | 5     | 0.005     |
| MF       | GO:0005737  | cytoplasm                                         | 18    | 0.033     |
| MF       | GO:0005198  | structural molecule activity                      | 18    | 1.40E-21  |
| MF       | GO:0005200  | structural constituent of cytoskeleton            | 9     | 1.31E-10  |

*Note:* P < 0.05 was considered significant.

Abbreviations: BP, biological process; CC, cellular component; GO, gene ontology; MF, molecular function.

Finally, we conducted the pie charts of 11 KRTs with ClueGO plugin of Cytoscape. The detailed functional notes and classification pie charts were listed in Fig. 8F. 44.44% terms belong to cornification, 22.22% to staphylococcus aureus infection, 11.11% to structural constituent of cytoskeleton, 11.11% to keratin filament, and 11.11% to intermediate filament cytoskeleton.

Relationship between the KRTs and clinical features in patients with LUSC. Pearson’s chi-squared test, which based on the TCGA database, indicated that the expression level of KRT13 was positively correlated with nodal metastasis (OR = 1.56 (1.07–2.27), P = 0.019), the expression level of KRT15 was positively correlated with primary tumor (OR = 33.78 (17.51–65.19), P = 0.000) and TNM stage (OR = 1.63 (1.03–2.60), P = 0.038), the expression level of KRT16 was positively correlated with nodal metastasis (OR = 56.67 (17.71-181.29), P = 0.000), the expression level of KRT17 was positively correlated with primary tumor (OR = 1.09 (1.07–2.67), P = 0.023) and nodal metastasis (OR = 38.24 (18.48–79.12), P = 0.000), and the expression level of KRT19 was positively correlated with nodal metastasis (OR = 1.58 (1.08–2.29), P = 0.017) (Table 4, Fig. 9).
## Table 4
Correlation between KRTs expression and the clinicopathologic features of LUSC patients

| Clinical characteristics | Total (N) | OR in KRT5 expression | P value | OR in KRT6A expression | P value | OR in KRT6B expression | P value | OR in KRT6C expression | P value |
|--------------------------|-----------|-----------------------|---------|------------------------|---------|------------------------|---------|------------------------|---------|
| Primary tumor (T1−2 vs. T3−4) | 501       | 0.95 (0.60–1.48)     | 0.813   | 1.29 (0.82–2.03)      | 0.261   | 1.23 (0.78–1.93)      | 0.371   | 1.36 (0.87–2.14)      | 0.176   |
| Lymph nodes (N0 vs. N1−3) | 485       | 1.26 (0.87–1.82)     | 0.231   | 1.09 (0.75–1.58)      | 0.660   | 1.01 (0.70–1.47)      | 0.951   | 1.05 (0.72–1.52)      | 0.802   |
| Distant metastasis (M0 vs. M1) | 410      | 1.34 (0.30–6.06)     | 0.703   | 0.75 (0.16–3.38)      | 0.703   | 1.34 (0.30–6.06)      | 0.703   | 0.75 (0.16–3.38)      | 0.703   |
| Stage (I-II vs. III-IV)  | 487       | 1.17 (0.74–1.85)     | 0.497   | 1.05 (0.66–1.66)      | 0.832   | 0.99 (0.63–1.57)      | 0.983   | 1.17 (0.74–1.85)      | 0.497   |

| OR in KRT13 expression | P value | OR in KRT14 expression | P value | OR in KRT15 expression | P value | OR in KRT16 expression | P value |
|------------------------|---------|------------------------|---------|------------------------|---------|------------------------|---------|
| Primary tumor (T1−2 vs. T3−4) | 501       | 1.05 (0.67–1.64)     | 0.836   | 1.23 (0.78–1.93)      | 0.731   | 33.78 (17.51–65.19)   | 0.000   | 1.29 (0.82–2.03)      | 0.261   |
| Lymph nodes (N0 vs. N1−3) | 485       | 1.56 (1.07–2.27)     | 0.019   | 0.94 (0.65–1.37)      | 0.750   | 1.13 (0.78–1.65)      | 0.529   | 56.67 (17.71–181.29)  | 0.000   |
| Distant metastasis (M0 vs. M1) | 410      | 0.75 (0.16–3.38)     | 0.703   | 0.75 (0.16–3.38)      | 0.703   | 2.54 (0.49–13.23)     | 0.253   | 0.39 (0.08–2.05)      | 0.253   |
| Stage (I-II vs. III-IV)  | 487       | 1.17 (0.74–1.85)     | 0.497   | 1.11 (0.70–1.75)      | 0.656   | 1.63 (1.03–2.60)      | 0.038   | 1.11 (0.70–1.75)      | 0.656   |

| OR in KRT17 expression | P value | OR in KRT19 expression | P value | OR in KRT23 expression | P value |
|------------------------|---------|------------------------|---------|------------------------|---------|
| Primary tumor (T1−2 vs. T3−4) | 501       | 1.09 (1.07–2.67)     | 0.023   | 1.52 (0.96–2.39)      | 0.070   | 1.16 (0.74–1.83)      | 0.506   |
| Lymph nodes (N0 vs. N1−3) | 485       | 38.24 (18.48–79.12)  | 0.000   | 1.58 (1.08–2.29)      | 0.017   | 0.88 (0.60–1.27)      | 0.486   |
| Distant metastasis (M0 vs. M1) | 410      | 1.34 (0.30–6.06)     | 0.703   | 0.75 (0.16–3.38)      | 0.703   | 2.54 (0.49–13.23)     | 0.253   |
| Stage (I-II vs. III-IV)  | 487       | 1.38 (0.87–2.19)     | 0.168   | 1.54 (0.97–2.46)      | 0.065   | 0.99 (0.63–1.57)      | 0.983   |

**Note:** P < 0.05 was considered significant.

**Abbreviations:** KRTs, keratins; LUSC, lung squamous cell carcinoma; OR, odd ratio.

---

## Discussion


Numerous studies have reported that the expression levels of KRTs in many tumors were dysregulated\textsuperscript{16–20}, and the roles of KRTs in occurrence, progression, and metastasis of multiple cancers has been partially confirmed\textsuperscript{36–39}. However, systematic bioinformatics analyses about the function of KRTs in LUSC are limited and have yet to be performed. This is the first time to systematically study the expression of KRTs in LUSC and their prognostic values. We hope our findings will help to provide new therapeutic measures and improve the prognosis of patients with LUSC.

KRTs family is a group of intermediate filament proteins which express in various types of epithelial cells\textsuperscript{40,41}. Previous studies have confirmed that KRTs played significant roles in maintaining the structural stability of epithelial cells and were involved in numerous biological activities, including epithelial cell signal transduction, stress response, tumor cell apoptosis, and tumor cell proliferation\textsuperscript{42,43}. Moreover, Nazarian et al.\textsuperscript{44} reported that the KRTs were dysregulated in a variety of tumor tissues, and played vital roles in invasion and metastasis of tumors. In summary, this kind of proteins has great potential in diagnosis, typing, and prognosis prediction of tumors. Therefore, the roles of KRTs as multifunctional regulators of epithelial tumorigenesis are worthy of further exploration. We used TCGA database to screen abnormally expressed KRTs between LUSC samples and adjacent normal lung samples, and found that 57 KRTs met our screening criteria. Subsequently, we used GEPIA database and HPA database to observe the transcriptional levels and protein levels of 57 KRTs. The results showed that the expression levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 in LUSC tissues were significantly higher than in adjacent normal lung tissues. Survival analyses showed that the high expression levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 were closely related to the poor OS and PFS of patients with LUSC. ROC curves indicated that the KRT5, KRT13, and KRT23 had potential diagnostic value in patients with LUSC.

KRT5 is usually specifically expressed in the basal layer of the epidermis, and its normal expression is essential for the protection of epithelial cells\textsuperscript{45}. Moll et al.\textsuperscript{19} reported that KRT5 was specifically expressed in squamous cell carcinoma of multiple tissues. For lung cancer, the expression level of KRT5 in LUSC tissues was significantly higher than in adjacent normal lung tissues\textsuperscript{20,46,47}, which was helpful for the differential diagnosis of LUSC and LUAD\textsuperscript{47}. In our study, TCGA database revealed that the expression level of KRT5 was higher in LUSC tissues than in adjacent normal lung tissues. By analyzing the Kaplan-Meier plotter database and TCGA database, we determined the prognostic value of KRT5 in patients with LUSC, and the results indicated that the high KRT5 expression was significantly associated with poor OS and PFS in patients with LUSC followed up for 200 months. In addition, the AUC of KRT5 was 60.753, which had a certain predictive effect on the prognosis of LUSC.

KRT6 is a group of keratins associated with stratified epithelium, including KRT6A, KRT6B, and KRT6C. Studies have reported that KRT6A, KRT6B, and KRT6C were involved in the epidermization of squamous epithelium, so their expression in LUSC increases sharply\textsuperscript{46–49}. The KRT6A is an important part of type II keratin family, and plays a considerable role in the epidermization of squamous epithelium\textsuperscript{50,51}. The overexpression of KRT6A in LUSC and LUAD can be used as a prognostic marker of lung cancer\textsuperscript{47,52,53}. In addition, studies have reported that KRT6A and KRT6B were related to the progression of renal carcinoma and breast cancer\textsuperscript{54,55}. The expression levels of KRT6A, KRT6B, and KRT6C in squamous cell carcinoma-like (SCCL) subtype urothelial carcinoma were significantly increased, and were closely related to poor prognosis\textsuperscript{56}. In our report, the expression levels of KRT6A, KRT6B, and KRT6C in LUSC tissues were higher than in normal lung tissues, which was consistent with the reported in the literature. However, there is no previous report on the relationship between KRT6 expression and tumor prognosis. Our research found that the high expression levels of KRT6A, KRT6B, and KRT6C were significantly associated with poor OS and PFS in patients with LUSC.

KRT13 encodes an intermediate cytoskeletal protein, which is usually expressed in the basal compartments of stratified epithelium\textsuperscript{57}. Previous studies have reported that KRT13 was expressed at different levels in different tumors, low KRT13 expression was found in oral squamous carcinomas\textsuperscript{58}, esophageal squamous cell carcinoma\textsuperscript{59}, bladder cancer\textsuperscript{60}, lymph node-positive uterine cervix cancer\textsuperscript{61}, and head and neck squamous cell carcinoma cell lines\textsuperscript{62}, by contrast, high KRT13
expression was detected in LUSC, colorectal cancer, gastric cancer, and tongue squamous cell carcinoma. The overexpression of KRT13 was associated with the squamous cell metaplasia of the upper airway tract epithelium, and was prominent among the gene expression markers for LUSC. Yang et al. found that the HOXC6 promotes lung cancer progression by upregulating KRT13 expression. In addition, KRT13 was implicated in urothelial and stem cell differentiation. In our report, we demonstrated that the expression level of KRT13 in LUSC tissues was higher than in normal lung tissues, the high KRT13 expression was correlated with poor prognosis in patients with LUSC, and the sensitivity of KRT13 was highest, which seemed consistent with the role of KRT13 as an oncogene and a diagnosis factor of LUSC.

KRT14 is an intermediate filament protein, which is usually expressed by basal epithelial progenitor cells located in epithelial niches of healthy adult tissues. Previous study indicated that KRT14 was not detected in normal tissues, was significantly increased in many tumor tissues, and was expressed in squamous cell carcinoma of different origins and degrees of differentiation. According to the IHC analysis of surgical specimens, the expression level of KRT14 in lung cancer tissues was significantly increased, especially in LUSC tissues, which indicated that KRT14 played an important role in the occurrence and development of LUSC. Chen et al. found that the high KRT14 expression could be helpful in the differential diagnosis of LUAD and LUSC. In addition, Huang et al. reported that KRT14 was not expressed in normal cervical tissue, but with the increase of the grade of cervical epithelial neoplasia, the expression level of KRT14 gradually increased, which indicated that KRT14 could be used as one of the indicators for the diagnosis of early cervical cancer. In our study, we demonstrated that the expression level of KRT14 in LUSC tissues was higher than in cancer-free lung tissues, and high KRT14 expression was related to poor OS and PFS in patients with LUSC.

KRT15 is a type I keratin, which is mainly expressed in basal keratinocytes of stratified epithelium and plays a significant role in maintaining cytoplasmic stability. Previous studies confirmed that the overexpression of KRT15 was involved in tumor formation and progression, including NSCLC, breast cancer, oral squamous neoplasms, urothelial cell carcinomas, and hepatocellular carcinoma. Sanchez-Palencia et al. reported that KRT15 expression level was higher in LUSC tissues than in cancer-free lung tissues, and may be a marker of LUSC. In addition, the high expression KRT15 was associated with poor prognosis of colorectal cancer and gastric cancer. In the present study, KRT15 was significantly overexpressed in LUSC tissues, and its high expression level was related to the poor prognosis in patients with LUSC.

KRT16 is an important part of type I cytoskeleton. KRT16 was reported to play a role in LUAD tumorigenicity via EMT, and increased KRT16 expression was associated with poor outcomes of tumors, including LUAD, oral squamous cell carcinoma, and metastatic breast cancer, which indicated KRT16 had an oncogenic role in tumors. In our report, we demonstrated that the expression level of KRT16 was higher in LUSC tissues, which was significantly related to poor OS and PFS in patients with LUSC.

KRT17 is a 48 kDa type I keratin. Extensive tissue screening results showed that KRT17 was low expressed in mature epithelial tissues, but regenerated and highly expressed in cancer tissues, and high expression of KRT17 indicated poor prognosis in NSCLC, which indicated that KRT17 can be a biomarker in predicting progression and poor prognosis in patients with LUSC. Chen et al. reported that KRT17 was highly expressed in LUSC, and in LUSC was significantly higher than in LUAD, suggesting that KRT17 may be a tumor marker for differentiating LUSC and LUAD subtypes of lung cancer. In addition, KRT17 was also highly expressed in breast cancer, cutaneous squamous cell carcinoma, cervical carcinoma, gastric carcinoma, and oral sprays, which was associated with poor prognosis of these tumors. In this report, we found that KRT17 was highly expression in LUSC tissues, which was correlated with poor OS and PFS of the patients with LUSC.

KRT19 is a member of the keratin family, which is functionally related to maintaining the structural integrity of epithelial cells, including bronchial epithelial cells. Yuan et al. discovered that compared with LUAD, KRT19 was upregulated in LUSC, and the high expression level of KRT19 was closely related to poor prognosis of LUSC. In addition, KRT19 has been
identified as a tumor-associated biomarker for a variety of tumors, including esophageal squamous cell carcinoma\(^{97}\), cervical carcinoma\(^{98}\), breast cancer\(^{99}\), colorectal cancer\(^{100}\), and hepatocellular carcinoma\(^{101}\). Moreover, the overexpression of KRT19 was associated with poor prognosis of hepatocellular carcinoma\(^{102}\) and breast cancer\(^{103}\). In the present study, KRT19 was significantly overexpressed in LUSC tissues, which was related to poor OS and PFS in LUSC patients.

KRT23 is an important member of the KRT gene family, and locates on human chromosome 17q21.2. Previous studies have shown that KRT23 is abnormally expressed in a variety of tumor tissues, including pancreatic carcinoma\(^{104}\), colorectal carcinoma\(^{105}\), and hepatocellular carcinoma\(^{106}\). In our report, we demonstrated that the expression level of KRT23 was higher in LUSC tissues than in normal lung tissues, and the high KRT23 expression was significantly correlated with poor prognosis in patients with LUSC. ROC curve analysis found that the specificity of KRT23 was 86.31, which had a certain predictive effect on the prognosis of LUSC.

In our study, the analysis of co-expression of KRTs found that KRTs were co-expressed in different degrees in LUSC, which was consistent with the finding of Travis, who found that KRT5, KRT6, KRT13, KRT14, KRT16, KRT17, and KRT19 were the gene expression markers for LUSC\(^{46}\).

Subsequently, we used cbioportal to detect the possible genetic alterations of KRTs in LUSC. The result indicated that mutation was the commonest gene alteration in KRTs. Previous studies have shown that the mutations of transcription factors were related to the occurrence and development of cancer, the abnormal expression of oncogenic transcription factors and chromatin regulatory factors were the internal reason for the tumorigenesis of cancer cells\(^{107}\). The abnormal function of these regulatory factors provided some clues for the mechanisms of abnormal gene expression\(^{108,109}\).

GO function and KEGG pathway analyses found that the KRTs and their neighbor genes were mainly enriched in epidermis development, keratinization, keratinocyte differentiation, and Rap1 signaling pathway. This was consistent with the knowledge that cytokeratin was abundant and stable in epithelial cells, and it had high specificity in tissues and cells, so it could be used as an ideal biological indicator to identify the origin of tumor and predict prognosis\(^{110}\). Some studies have reported that keratinization of tumors was closely related to the poor prognosis of tumor patients\(^{111,112}\), suggesting the importance of keratinization in neoplastic prognosis\(^{111}\). Studies on Rap1 signaling pathway have reported that abnormality of Rap1 signaling pathway was highly associated with the occurrence and development of some tumors, including pancreatic cancer, melanoma, and human chronic myelogenous leukemia\(^{113–115}\).

Finally, we analyzed the relationship between the expression level of KRTs and the clinical characteristics of patients with LUSC. It turned out that the expression levels of KRT13, KRT16, KRT17, and KRT19 were positively correlated with nodal metastasis, the expression levels of KRT15 and KRT17 were positively related to primary tumor, and the expression level of KRT15 was positively correlated with TNM stage. There were some studies on the relationship between KRTs and clinical features of tumors. Liu et al.\(^{85}\) reported that KRT16 was upregulated in LUAD and the highly KRT16 expression was related to nodal metastasis. Huang et al.\(^{36}\) suggested that high expression of KRT16 indicates the poorer pathological differentiation, advanced stages, and increased nodal metastasis of OSCC patients. In addition, the expression level of KRT16 was significantly increased in the process of nodal metastasis in primary breast cancer\(^{17}\). Wang et al.\(^{88}\) found that the highly expressed KRT17 was associated with poor differentiation and nodal metastasis in patients with NSCLC. Hu et al.\(^{116}\) found that the expression level of KRT17 was significantly associated with the tumor size, depth of invasion, nodal metastasis, and TNM stage in patients with gastric cancer. Nordgard et al.\(^{117}\) confirmed that KRT19 can be used as a molecular biomarker of disseminated tumor cells in regional lymph nodes from NSCLC patients.

In this study, we comprehensively analyzed the effects of KRTs on LUSC, and provided a thorough understanding of the heterogeneity and complexity of the molecular mechanisms of LUSC. Our results indicated that the increased expression of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 in LUSC tissues might play an important role in LUSC oncogenesis and progression, and could also serve as molecular markers to identify high-risk group
of LUSC patients. In addition, our findings suggested that KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 were potential therapeutic targets for LUSC, and were potential prognostic markers for patients with LUSC.

Conclusions

In brief, our study is the first to demonstrate that the expression levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 are elevated in LUSC tissues compared with normal lung tissues, and the high expression levels of KRT5, KRT6A, KRT6B, KRT6C, KRT13, KRT14, KRT15, KRT16, KRT17, KRT19, and KRT23 predict poor prognosis. These novel findings not only shed light on the molecular alterations in LUSC but also provide the foundation for further research in this area.

Declarations

Data Availability Statement

High-throughput gene expression data of LUSC tissues and normal lung tissues were extracted from the Cancer Genome Atlas (TCGA) data portal (https://tcga-data.nci.nih.gov/tcga). These RNA-seq data (HTSeq-count) from Illumina HiSeq RNASeq platform consisted of 502 LUSC samples and 49 adjacent non-cancerous lung samples, and were achieved from the publicly available Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/).

Author Contribution Statement

Qiangqiang Zheng designed this study, Qiangqiang Zheng analyzed the data. Qiangqiang Zheng wrote the manuscript. All authors reviewed the manuscript.

Conflict of Interest Statement

The authors declare no competing financial interests and competing non-financial interest.

Ethics Statement

Not applicable.

Acknowledgements

Not applicable.

References

1. Chen, W. et al. Cancer statistics in China, 2015. CA-A Cancer J Clin. 66, 115. https://doi.org/10.3322/caac.21338 (2016).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. CA-A Cancer J Clin. 70, 7-30. https://doi.org/10.3322/caac.21590 (2020).
3. Allemani, C. et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet. 391, 1023-1075. https://doi.org/10.1016/S0140-6736(17)33326-3 (2018).
4. Chang, J. T., Lee, Y. M. & Huang, R.S. The impact of the Cancer Genome Atlas on lung cancer. Transl Res. 166, 568-585. https://doi.org/10.1016/j.trsl.2015.08.001 (2015).
5. Kenfield, S.A. et al. Comparison of aspects of smoking among the four histological types of lung cancer. Tob Control. 17, 198-204. https://doi.org/10.1136/tc.2007.022582 (2008).
6. Abbas, A., Aster, J. & Kumar, V. *Robbins Basic Pathology, 9th Ed* (2012).

7. Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* **489**, 519-525. doi:10.1038/nature11404 (2012).

8. Collisson, E.A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* **511**, 543-550. doi:10.1038/nature13385 (2014).

9. Rekhtman, N. *et al.* Clarifying the spectrum of driver oncogene mutations in biomarker-verified squamous carcinoma of lung: lack of EGFR/KRAS and presence of PIK3CA/AKT1 mutations. *Clin Cancer Res.* **18**, 1167-1176. doi:10.1158/1078-0432.CCR-11-2109 (2012).

10. Bleeker, F. E. *et al.* AKT1(E17K) in human solid tumours. *Oncogene.* **27**, 5648-5650. doi:10.1038/onc.2008.170 (2008).

11. Brose, M.S. *et al.* BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res.* **62**, 6997-7000 (2002).

12. Bean, J. *et al.* MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. *Proc Natl Acad Sci U S A.* **104**, 20932-20937. doi:10.1073/pnas.0710370104 (2007).

13. Jin, G. *et al.* PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. *Lung Cancer.* **69**, 279-283. doi:10.1016/j.lungcan.2009.11.012 (2010).

14. Koster, M. I. Building models for keratin disorders. *J Invest Dermatol.* **132**, 1324-1326. doi:10.1038/jid.2012.58 (2012).

15. Kerns, M. *et al.* Differential modulation of keratin expression by sulforaphane occurs via Nrf2-dependent and -independent pathways in skin epithelia. *Mol Biol Cell.* **21**, 4068-4075. doi:10.1091/mbc.E10-02-0153 (2010).

16. Fujioka, M. *et al.* Dimethylarsinic acid (DMA) enhanced lung carcinogenesis via histone H3K9 modification in a transplacental mouse model. *Arch Toxicol.* **94**, 927-937. doi:10.1007/s00204-020-02665-x (2020).

17. Joosse, S. A. *et al.* Changes in keratin expression during metastatic progression of breast cancer: impact on the detection of circulating tumor cells. *Clin Cancer Res.* **18**, 993-1003. doi:10.1158/1078-0432.CCR-11-2100 (2012).

18. Misiorek, J. O. *et al.* Keratin 8-deletion induced colitis predisposes to murine colorectal cancer enforced by the inammasome and IL-22 pathway. *Carcinogenesis.* **37**, 777-786. doi:10.1093/carcin/bgw063 (2016).

19. Moll, R., Divo, M. & Langbein, L. The human keratins: biology and pathology. *Histochem Cell Biol.* **129**, 705-733. doi:10.1007/s00418-008-0435-6 (2008).

20. Karantza, V. Keratins in health and cancer: more than mere epithelial cell markers. *Oncogene.* **30**, 127-138. doi:10.1038/onc.2010.456 (2011).

21. Zhu, M. *et al.* Reconstructing differentially co-expressed gene modules and regulatory networks of soybean cells. *BMC Genomics.* **13**, 437. doi:10.1186/1471-2164-13-437 (2012).

22. Bejjani, B. A. & Shaffer, L. G. Clinical utility of contemporary molecular cytogenetics. *Annu Rev Genomics Hum Genet.* **9**, 71-86. doi:10.1146/annurev.genom.9.081307.164207 (2008).

23. Zhang, C. *et al.* Profiling alternatively spliced mRNA isoforms for prostate cancer classification. *BMC Bioinformatics.* **7**, 202. doi:10.1186/1471-2105-7-202 (2006).

24. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. doi:10.1093/nar/gkv007 (2015).

25. Tang, Z. *et al.* GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556-W560. doi:10.1093/nar/gkz430 (2019).

26. Pontén, F., Jirström, K. & Uhlen, M. The Human Protein Atlas--a tool for pathology. *J Pathol.* **216**, 387-393. doi:10.1002/path.2440 (2008).
27. Hou, G. X. et al. Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan-Meier plotter. *PLoS One* **12**, e0174515. [https://doi.org/10.1371/journal.pone.0174515](https://doi.org/10.1371/journal.pone.0174515) (2017).

28. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77. [https://doi.org/10.1186/1471-2105-12-77](https://doi.org/10.1186/1471-2105-12-77) (2011).

29. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, p1. [https://doi.org/10.1126/scisignal.2004088](https://doi.org/10.1126/scisignal.2004088) (2013).

30. Ogata, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **27**, 29-34. [https://doi.org/10.1093/nar/27.1.29](https://doi.org/10.1093/nar/27.1.29) (1999).

31. Kohl, M., Wiese, S. & Warscheid, B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol* **696**, 291-303. [https://doi.org/10.1007/978-1-60761-987-1_18](https://doi.org/10.1007/978-1-60761-987-1_18) (2011).

32. Jiao, X. et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **28**, 1805-1806. [https://doi.org/10.1093/bioinformatics/bts251](https://doi.org/10.1093/bioinformatics/bts251) (2012).

33. Raman, P. et al. Pancreatic cancer survival analysis defines a signature that predicts outcome. *PLoS One* **13**, e0201751. [https://doi.org/10.1371/journal.pone.0201751](https://doi.org/10.1371/journal.pone.0201751) (2018).

34. van de Rijn, M. et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* **161**, 1991-1996. [https://doi.org/10.1016/S0002-9440(10)64476-8](https://doi.org/10.1016/S0002-9440(10)64476-8) (2002).

35. Guo, L. et al. Cytokeratin-20 and Survivin-Expressing Circulating Tumor Cells Predict Survival in Metastatic Colorectal Cancer Patients by a Combined Immunomagnetic qRT-PCR Approach. *Mol Cancer Ther* **14**, 2401-2408. [https://doi.org/10.1158/1535-7163.MCT-15-0359](https://doi.org/10.1158/1535-7163.MCT-15-0359) (2015).

36. Nazarian, R. M. et al. Cytokeratin 17: an adjunctive marker of invasion in squamous neoplastic lesions of the anus. *Am J Surg Pathol* **38**, 78-85. [https://doi.org/10.1097/PAS.00000000000000111](https://doi.org/10.1097/PAS.00000000000000111) (2014).

37. Guo, L. et al. A novel heterozygous nonsense mutation of keratin 5 in a Chinese family with Dowling-Degos disease. *J Eur Acad Dermatol Venereol* **26**, 908-910. [https://doi.org/10.1111/j.1468-3083.2011.04115.x](https://doi.org/10.1111/j.1468-3083.2011.04115.x) (2012).

38. Travis, D. et al. WHO classification. Tumors of the lung, pleura, thymus and heart. *Albany, NY: WHO Publication Center* (2004).
47. Chen, Y. et al. The diagnostic value of cytokeratin 5/6, 14, and 17, and 18 expression in human non-small cell lung cancer. *Oncology*. **80**, 333-340. https://doi.org/10.1159/000329098 (2011).

48. Hawthorn, L. et al. Characterization of cell-type specific profiles in tissues and isolated cells from squamous cell carcinomas of the lung. *Lung Cancer*. **53**, 129-142. https://doi.org/10.1016/j.lungcan.2006.04.015 (2006).

49. Fuji, T. et al. A preliminary transcriptome map of non-small cell lung cancer. *Cancer Res*. **62**, 3340-3346 (2002).

50. Chen, C. & Shan, H. Keratin 6A gene silencing suppresses cell invasion and metastasis of nasopharyngeal carcinoma via the β-catenin cascade. *Mol Med Rep*. **19**, 3477-3484. https://doi.org/10.3892/mmr.2019.10055 (2019).

51. Forrest, C. E. et al. Pachyonychia Congenita: A Spectrum of KRT6a Mutations in Australian Patients. *Pediatr Dermatol*. **33**, 337-342. https://doi.org/10.1111/pde.12841 (2016).

52. Chang, H. H., Dreyfuss, J. M. & Ramoni, M. F. A transcriptional signature characterizes lung cancer subtypes. *Cancer*. **117**, 353-360. https://doi.org/10.1002/cncr.25592 (2011).

53. Yang, B. et al. KRT6A Promotes EMT and Cancer Stem Cell Transformation in Lung Adenocarcinoma. *Technol Cancer Res Treat*. **19**, 1533033820921248. https://doi.org/10.1177/1533033820921248 (2020).

54. Hu, J. et al. KRT6 interacting with notch1 contributes to progression of renal cell carcinoma, and aliskiren inhibits renal carcinoma cell lines proliferation in vitro. *Int J Clin Exp Pathol*. **8**, 9182-9188 (2015).

55. Bu, W. et al. Keratin 6a marks mammary bipotential progenitor cells that can give rise to a unique tumor model resembling human normal-like breast cancer. *Oncogene*. **30**, 4399-4409. https://doi.org/10.1038/onc.2011.147 (2011).

56. Sjödahl, G. et al. A molecular taxonomy for urothelial carcinoma. *Clin Cancer Res*. **18**, 3377-3386. https://doi.org/10.1158/1078-0432.CCR-12-0077-T (2012).

57. Moll, R. et al. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell*. **31**, 11-24. https://doi.org/10.1016/0092-8674(82)90400-7 (1982).

58. Bloor, B. K., Seddon, S. V. & Morgan, P. R. Gene expression of differentiation-specific keratins in oral epithelial dysplasia and squamous cell carcinoma. *Oral Oncol*. **37**, 251-261. https://doi.org/10.1016/0928-4384(82)90400-7 (2001).

59. Takashima, M. et al. Reduced expression of cytokeratin 4 and 13 is a valuable marker for histologic grading of esophageal squamous intraepithelial neoplasia. *J Med Dent Sci*. **59**, 17-28 (2012).

60. Marsit, C. J. et al. Identification of methylated genes associated with aggressive bladder cancer. *PLoS One*. **5**, e12334. https://doi.org/10.1371/journal.pone.0012334 (2010).

61. van Bommel, P. F. et al. Expression of cytokeratin 10, 13, and involucrin as prognostic factors in low stage squamous cell carcinoma of the uterine cervix. *Cancer*. **74**, 2314-2320. https://doi.org/10.1002/1097-0142(19941015)74:8<2314::aid-cncr28207408163.0.co;2-1 (1994).

62. Baker, H. et al. Proteome-wide analysis of head and neck squamous cell carcinomas using laser-capture microdissection and tandem mass spectrometry. *Oral Oncol*. **41**, 183-199. https://doi.org/10.1016/j.oraloncology.2004.08.009 (2005).

63. Madoz-Gúrpide, J. et al. A proteomics analysis of cell signaling alterations in colorectal cancer[J]. *Mol Cell Proteomics*. **6**, 2150-2164. https://doi.org/10.1074/mcp.M700006-MCP200 (2007).

64. Lian, F. et al. Farnesoid X receptor protects human and murine gastric epithelial cells against inflammation-induced damage. *Biochem J*. **438**, 315-323. https://doi.org/10.1042/BJ20102096 (2011).

65. Yanagawa, T. et al. Loss of cytokeratin 13 expression in squamous cell carcinoma of the tongue is a possible sign for local recurrence. *J Exp Clin Canc Res*. **26**, 215-220 (2007).

66. Bolton, S.J. et al. Characterisation of the proximal airway squamous metaplasia induced by chronic tobacco smoke exposure in spontaneously hypertensive rats. *Respir Res*. **10**, 118. https://doi.org/10.1186/1465-9921-10-118 (2009).

67. Schlage, W.K. et al. Cytokeratin expression patterns in the rat respiratory tract as markers of epithelial differentiation in inhalation toxicology. II. Changes in cytokeratin expression patterns following 8-day exposure to room-aged cigarette sidestream smoke. *Toxicologic Pathology*. **26**, 344-360. https://doi.org/10.1177/019262339802600308 (1998).
68. Yang, Y. et al. Evidence for an oncogenic role of HOXC6 in human non-small cell lung cancer. PeerJ. 7, e6629. https://doi.org/10.7717/peerj.6629 (2019).

69. Southgate, J. et al. Differentiation potential of urothelium from patients with benign bladder dysfunction. BJU Int. 99, 1506-1516. https://doi.org/10.1111/j.1464-410X.2007.06795.x (2007).

70. Jankowski, M. et al. Novel KRT14 mutation causing epidermolysis bullosa simplex with variable phenotype. Exp Dermatol. 23, 684-687. https://doi.org/10.1111/exd.12478 (2014).

71. Poovorawan, N. et al. Analysis of KRT14 and SFTPB expression by immunohistochemistry method in squamous cell carcinoma of lung and head-neck cancer tissue. Journal of Clinical Oncology. 32, e22127. https://doi.org/10.1200/jco.2014.32.15_suppl.e22127 (2014).

72. Huang, L. et al. Identification of a 7-gene signature that predicts relapse and survival for early stage patients with cervical carcinoma. Med Oncol. 29, 2911-2918. https://doi.org/10.1007/s12032-012-0166-3 (2012).

73. Chong, L.Y. et al. Keratin 15, transcobalamin I and homeobox gene Hox-B13 expression in breast phyllodes tumors: novel markers in biological classification. Breast Cancer Res Treat. 132, 143-151. https://doi.org/10.1007/s10549-011-1555-6 (2012).

74. Mahalingam, M., Srivastava, A. & Hoang, M. P. Expression of stem-cell markers (cytokeratin 15 and nestin) in primary adnexal neoplasms-clues to etiopathogenesis. Am J Dermatopathol. 32, 774-779. https://doi.org/10.1097/DA.0b013e3181d9f8c (2010).

75. Gómez-Morales, M. et al. Differential immunohistochemical localization of desmosomal plaque-related proteins in non-small cell lung cancer. Histopathology. 63, 103-113. https://doi.org/10.1111/his.12126 (2013).

76. Sanchez-Palencia, A. et al. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. Int J Cancer. 129, 355-364. https://doi.org/10.1002/ijc.25704 (2011).

77. Cimino, D. et al. Identification of new genes associated with breast cancer progression by gene expression analysis of predefined sets of neoplastic tissues. Int J Cancer. 123, 1327-1338. https://doi.org/10.1002/ijc.23660 (2008).

78. Gelfand, R. et al. Long-term exposure of MCF-7 breast cancer cells to ethanol stimulates oncogenic features. Int J Oncol. 50, 49-65. https://doi.org/10.3892/ijo.2016.3800 (2017).

79. Khanom, R. et al. Expression of basal cell keratin 15 and keratin 19 in oral squamous neoplasms represents diverse pathophysiology. Histol Histopathol. 27, 949-959. https://doi.org/10.14670/HH-27.949 (2012).

80. Tai, G. et al. Cytokeratin 15 marks basal epithelia in developing ureters and is upregulated in a subset of urothelial cell carcinomas. PLoS One. 8, e81167. https://doi.org/10.1371/journal.pone.0081167 (2013).

81. Zekri, A. N. et al. Gene expression profiling of circulating CD133+ cells of hepatocellular carcinoma patients associated with HCV infection. J Egypt Natl Canc Inst. 29, 19-24. https://doi.org/10.1016/j.jnci.2016.12.002 (2017).

82. Sanchez-Palencia, A. et al. Gene expression profiling reveals novel biomarkers in non small cell lung cancer. Int J Cancer. 129, 355-364. https://doi.org/10.1002/ijc.25704 (2011).

83. Rao, X. et al. KRT15 overexpression predicts poor prognosis in colorectal cancer. Neoplasma. 67, 410-414. https://doi.org/10.4149/neop_2019_190531N475 (2020).

84. Zhang, C. et al. KRT15, INHBA, MATN3, and AGT are aberrantly methylated and differentially expressed in gastric cancer and associated with prognosis. Pathol Res Pract. 215, 893-899. https://doi.org/10.1016/j.prp.2019.01.034 (2019).

85. Liu, Y. H. et al. TFAP2A Induced KRT16 as an Oncogene in Lung Adenocarcinoma via EMT. Int J Biol Sci. 15, 1419-1428. https://doi.org/10.7150/ijbs.34076 (2019).

86. Polioudaki, H. et al. Variable expression levels of keratin and vimentin reveal differential EMT status of circulating tumor cells and correlation with clinical characteristics and outcome of patients with metastatic breast cancer. BMC Cancer. 15, 399. https://doi.org/10.1186/s12885-015-1386-7 (2015).

87. Troyanovsky, S. M., Leube, R. E. & Franke, W. W. Characterization of the human gene encoding cytokeratin 17 and its expression pattern. Eur J Cell Biol. 59, 127 (1992).
88. Wang, Z. et al. Overexpression of KRT17 promotes proliferation and invasion of non-small cell lung cancer and indicates poor prognosis. Cancer Manag Res. 11, 7485-7497. https://doi.org/10.2147/CMAR.S218926 (2019).

89. Matt, V. D. R. et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol. 161, 1991-1996. https://doi.org/10.1016/S0002-9440(10)64476-8 (2002).

90. SoRlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 98, 10869-10874. https://doi.org/10.1073/pnas.191367098 (2001).

91. Moll, R., Divo, M. & Langbein, L. J. H. The human keratins: biology and pathology. Cell Biol. 129, 705-733. https://doi.org/10.1007/s00418-008-0435-6 (2008).

92. Escobar-Hoyos, L. F. et al. Keratin 17 in premalignant and malignant squamous lesions of the cervix: proteomic discovery and immunohistochemical validation as a diagnostic and prognostic biomarker. Mod Pathol. 27, 621-630. https://doi.org/10.1038/modpathol.2013.166 (2014).

93. Hu, H. et al. Keratin 17 promotes tumor growth and is associated with poor prognosis in gastric cancer. J Cancer. 9, 346-357. https://doi.org/10.7150/jca.19838 (2018).

94. Mikami, Y. et al. GLI-mediated Keratin 17 expression promotes tumor cell growth through the anti-apoptotic function in oral squamous cell carcinomas. J Cancer Res Clin Oncol. 143, 1381-1393. https://doi.org/10.1007/s00432-017-2398-2 (2017).

95. Saha, S. K. et al. Cytokeratin 19 (KRT19) has a Role in the Reprogramming of Cancer Stem Cell-like Cells to less aggressive and more drug-sensitive cells. Int J Mol Sci. 19, 1423. https://doi.org/10.3390/ijms19051423 (2018).

96. Yuan, X. et al. Prognostic significance of KRT19 in Lung Squamous Cancer. J Cancer. 12, 1240-1248. https://doi.org/10.7150/jca.51179 (2021).

97. Zhang, J. et al. Using proteomic approach to identify tumor-associated proteins as biomarkers in human esophageal squamous cell carcinoma. J Proteome Res. 10, 2863-2872. https://doi.org/10.1021/pr200141c (2011).

98. Okamoto, S. et al. Detection of sentinel lymph node metastases in cervical cancer: assessment of KRT19 mRNA in the one-step nucleic acid amplification (OSNA) method. Gynecol Oncol. 130, 530-536. https://doi.org/10.1016/j.ygyno.2013.06.027 (2013).

99. Wang, X. M. et al. KRT19 and CEACAM5 mRNA-marked circulated tumor cells indicate unfavorable prognosis of breast cancer patients. Breast Cancer Res Treat. 174, 375-385. https://doi.org/10.1007/s10549-018-05069-9 (2019).

100. Mostert, B. et al. mRNA expression profiles in circulating tumor cells of metastatic colorectal cancer patients. Mol Oncol. 9, 920-932. https://doi.org/10.1016/j.molonc.2015.01.001 (2015).

101. Zeng, S. S. et al. The transcription factor SALL4 regulates stemness of EpCAM-positive hepatocellular carcinoma. J Hepatol. 60, 127-134. https://doi.org/10.1016/j.jhep.2013.08.024 (2014).

102. Nault, J. C. et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. Gastroenterology. 145, 176-187. https://doi.org/10.1053/j.gastro.2013.03.051 (2013).

103. Ignatiadis, M. et al. Different prognostic value of cytokeratin-19 mRNA positive circulating tumor cells according to estrogen receptor and HER2 status in early-stage breast cancer. J Clin Oncol. 25, 5194-5202. https://doi.org/10.1200/JCO.2007.11.7762 (2007).

104. Zhang, J. S. et al. Keratin 23 (K23), a novel acidic keratin, is highly induced by histone deacetylase inhibitors during differentiation of pancreatic cancer cells. Genes Chromosomes Cancer. 30, 123-135 (2001).

105. Zhang, N. et al. Keratin 23 promotes telomerase reverse transcriptase expression and human colorectal cancer growth. Cell Death Dis. 8, e2961. https://doi.org/10.1038/cddis.2017.339 (2017).

106. Wang, K. et al. Identification of tumor-associated antigens by using SEREX in hepatocellular carcinoma. Cancer Lett. 281, 144-150. https://doi.org/10.1016/j.canlet.2009.02.037 (2009).

107. Sanda, T. et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell. 22, 209-221. https://doi.org/10.1016/j.ccr.2012.06.007 (2012).
108. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell*. **150**, 12-27. https://doi.org/10.1016/j.cell.2012.06.013 (2012).

109. Hargreaves, D. C. & Crabtree, G. R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* **21**, 396-420. https://doi.org/10.1038/cr.2011.32 (2011).

110. Kallen, M. E., Denicola, M. M. & Said, J. Coexpression of cytokeratin and B cell markers—a rare finding with new implications. *Journal of Hematopathology*. **8**, 1-3. https://doi.org/10.1007/s12308-015-0243-6 (2015).

111. Reddy, S. P. *et al.* Prognostic significance of keratinization in nasopharyngeal carcinoma. *Am J Otolaryngol*. **16**, 103-108. https://doi.org/10.1016/0196-0709(95)90040-3 (1995).

112. Jung, H. M. *et al.* Keratinization-associated miR-7 and miR-21 regulate tumor suppressor reversion-inducing cysteine-rich protein with kazal motifs (RECK) in oral cancer. *J Biol Chem*. **287**, 29261-29272. https://doi.org/10.1074/jbc.M112.366518 (2012).

113. Zhang, L. *et al.* Identification of a putative tumor suppressor gene Rap1GAP in pancreatic cancer. *Cancer Res*. **66**, 898-906. https://doi.org/10.1158/0008-5472.CAN-05-3025 (2006).

114. Gao, L. *et al.* Ras-associated protein-1 regulates extracellular signal-regulated kinase activation and migration in melanoma cells: two processes important to melanoma tumorigenesis and metastasis. *Cancer Res*. **66**, 7880-7888. https://doi.org/10.1158/0008-5472.CAN-06-0254 (2006).

115. Ishida, D. *et al.* Myeloproliferative stem cell disorders by deregulated Rap1 activation in SPA-1-deficient mice. *Cancer Cell*. **4**, 55-65. https://doi.org/10.1016/s1535-6108(03)00163-6 (2003).

116. Hu, H. *et al.* Keratin17 Promotes Tumor Growth and is Associated with Poor Prognosis in Gastric Cancer. *J Cancer*. **9**, 346-357. https://doi.org/10.7150/jca.19838 (2018).

117. Nordgard, O. *et al.* Novel molecular tumor cell markers in regional lymph nodes and blood samples from patients undergoing surgery for non-small cell lung cancer. *PLoS One*. **8**, e62153. https://doi.org/10.1371/journal.pone.0062153 (2013).

**Figures**
Figure 1

Heatmap and Volcanoplot of the distribution of KRTs between LUSC tissues and normal lung tissues. (A) Heatmap of 57 KRTs between LUSC tissues and normal lung tissues. The upregulated and downregulated genes were shown in red and green, respectively. (B) Volcanoplot of 57 KRTs between LUSC tissues and normal lung tissues. The upregulated genes were shown in red.
Figure 2

Transcriptional expression of distinct KRTs family members. (A) scatter diagram. (B) box plot (*P < 0.05).
Figure 3

The IHC-based protein expression of KRTs in LUSC tissues and normal lung tissues. All the IHC staining images were obtained from the HPA database.
Figure 4

The OS of LUSC patients with high or low expression of KRTs. The prognostic significance of the KRTs in patients with LUSC, according to the (A) Kaplan-Meier plotter database and (B) TCGA database. The red curve represents the survival curve of LUSC patients with high gene expression, the black curve represents the survival curve of LUSC patients with low gene expression.
Figure 5

ROC curves for the KRTs in LUSC. Red represents sensitive curve, blue indicates identify line. The X axis shows false positive rate, presented as “100-Specificity (%)”. The Y axis indicates true positive rate, shown as “Sensitivity (%).” These curves were provided by ‘pROC’ package of RStudio.

Figure 6

The genetic alterations of KRTs in LUSC. (A) The alteration frequency of KRTs. (B) Mutations of every KRT gene. Brown represents inframe mutation, green represents missense mutation, orange represents splice mutation, dark grey represents truncating mutation, red represents amplification, and blue represents deep deletion.
Figure 7

Co-expression of different KRTs in LUSC (RStudio).
Figure 8

The PPI network and enrichment analyses of KRTs and genes significantly associated with KRT alterations. (A) PPI network of KRTs and genes significantly associated with KRT alterations. (B) GO analysis of KRTs and genes significantly associated with KRT alterations. (C) KEGG functional enrichment analysis of KRTs and genes significantly associated with KRT alterations. (D) MCODE identification of the most significantly enriched module. The circle represents upregulated genes, and the triangle represents downregulated genes. (E) GO analysis of the most significantly enriched module. (F) ClueGO analysis of 11 KRTs.
Figure 9

Forest plots of OR for the association between KRTs expression and clinicopathological features in patients with LUSC.