Original Article

Exploration and validation of hub genes in lung adenocarcinoma based on bioinformatics analysis

Changyi Zeng[1,2#^], You Zhou[1#^], Wanqing Ye[1^], Zihan Fang[1^], Ke Wang[1,2^]

[1]Department of Preventive Medicine, Medical College, Hubei University of Arts and Science, Xiangyang, China; [2]Research Centre for Evidence-Based and Translational Medicine, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, China

Contributions: (I) Conception and design: C Zeng, Y Zhou, K Wang; (II) Administrative support: K Wang; (III) Provision of study materials or patients: C Zeng, Y Zhou, Z Fang, K Wang; (IV) Collection and assembly of data: C Zeng, Z Fang; (V) Data analysis and interpretation: C Zeng, Z Fang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Ke Wang. Department of Preventive Medicine, Medical College, Hubei University of Arts and Science, 296 Longzhong Road, Xiangyang 441053, China. Email: mqlhome76@163.com.

Background: Genomic abnormality is a crucial factor for lung cancer development. This study used bioinformatics analysis to explore the hub genes involved in lung adenocarcinoma.

Methods: The GeneCards, Comparative Toxicogenomics Database (CTD), and DISEASES databases were used to screen the genes associated with lung adenocarcinoma. The hub genes were then identified using WebGestalt. The Cancer Genome Atlas (TCGA), UALCAN, and the Human Protein Atlas (HPA) were used to validate the expression of hub genes. The predictive effects of hub genes on the risk of lung adenocarcinoma were evaluated using receiver operating characteristic (ROC) curve analysis. The Tumor-Immune System Interaction Database (TISIDB) was used to estimate the correlation between hub genes and immune infiltration.

Results: A total of 21 genes were defined as common genes associated with lung adenocarcinoma, and from these, *AKT1*, *CD44*, and *CDKN2A* were identified as hub genes. Significant differences in the hub gene mRNA and protein expression were observed between lung adenocarcinoma samples and normal samples derived from the TCGA and UALCAN databases. The area under the ROC curve (AUC) for *AKT1*, *CD44*, and *CDKN2A* in predicting lung adenocarcinoma risk was 0.847, 0.880, and 0.805, respectively, with sensitivity of 89.8%, 93.2%, and 94.9%, respectively. TISIDB analysis indicated that *AKT1*, *CD44*, and *CDKN2A* expression had a strong relationship with immune infiltration in lung adenocarcinoma.

Conclusions: These hub genes, *AKT1*, *CD44*, and *CDKN2A*, may represent tumor biomarkers that may contribute to the understanding, diagnosis, and treatment of lung adenocarcinoma.

Keywords: Lung adenocarcinoma; hub gene; bioinformatics analysis; immune infiltration.

Submitted Aug 29, 2022. Accepted for publication Oct 12, 2022.
doi: 10.21037/tcr-22-2225
View this article at: https://dx.doi.org/10.21037/tcr-22-2225

^ ORCID: Changyi Zeng, 0000-0002-0957-2553; You Zhou, 0000-0002-1669-9173; Wanqing Ye, 0000-0001-6098-1232; Zihan Fang, 0000-0002-3666-5413; Ke Wang, 0000-0003-0829-8510.

Introduction

Lung cancer is one of the most common carcinomas worldwide, with the highest rate of mortality and morbidity in both men and women (1). There are several categories of lung cancer according to pathological features, and non-small cell lung cancer (NSCLC), including lung adenocarcinoma and lung squamous carcinoma, represents almost 90% of all lung cancers (2). In the past few decades, the incidence of lung adenocarcinoma has surpassed that of lung squamous carcinoma to become the most common clinical type of NSCLC (3,4). Despite the development of therapeutic treatments for lung cancer patients, the 5-year relative survival rate remains unsatisfactory (5), with a reported 30% survival rate in primary lung cancer patients and less than 10% survival in metastatic lung cancer patients (6). Consequently, the development of new and more effective treatment methods for lung cancer is urgently needed.

Increasingly, studies have indicated that the biological mechanisms in lung cancer are caused by multiple factors working synergistically (7,8). Among them, genomic abnormalities are a crucial factor that cannot be ignored, especially in lung adenocarcinoma (9). Indeed, a deep sequencing study has demonstrated complex and wide heterogeneity in the lung cancer genome (10). With an increasing number of studies showing that immune infiltration is closely related to the development of lung adenocarcinoma, the direction for lung cancer treatment has changed and is now focused on individualized therapy according to the identification of immune-related biomarkers with obvious genetic abnormalities (11,12). With these discoveries, targeted therapy has been widely used in the clinic and is now defined as an important standard for clinical treatment in lung cancer patients (13). Consequently, the survival rate of some lung cancer patients has been moderately but steadily improved. Therefore, exploring new tumor biomarkers is particularly important for individualized therapy of lung adenocarcinoma patients.

Bioinformatics analysis is a fast and useful method for examining how all the molecular discoveries relevant to one patient can be applied to others (14). Generalization and summarization of key biological information based on comprehensive multisource data and biological networks can provide certain scientific evidence to support early diagnosis and individual medicine of disease. With bioinformatics research, molecular and genetic information of diseases can be well connected and further evaluated to understand the etiology of diseases, revealing the importance and complexity of multilevel research in disease progression, diagnosis, and individualized therapy.

This current study integrated the available bioinformatics information related to lung adenocarcinoma to screen and identify a group of genes associated with lung adenocarcinoma based on three comprehensive databases. These genes were validated and evaluated in terms of different expression, clinical application value and the underlying biological mechanism. This data will provide useful information for understanding the pathological mechanisms and diagnosis of lung adenocarcinoma, and provide certain scientific basis for individualized treatment in lung adenocarcinoma. We present the following article in accordance with the STREGA reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-2225/rc).

Methods

Public databases for gene screening

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Public databases were analyzed to screen the genes associated with lung adenocarcinoma. The following databases were used: (I) GeneCards (https://www.genecards.org/), which provides comprehensive information related to human genes, RNA, protein, and functional information from a wide variety of related websites (15); (II) Comparative Toxicogenomics Database (CTD, http://ctdbase.org/), which includes genomic, environmental, as well as chemical factors that may play major roles in human health, and assesses the relationship between disease pathogenesis and environmental factors (16); and (III) DISEASES (https://diseases.jensenlab.org/Search), which is a large and comprehensive database for disease-gene association studies, and information involving the genomic data of cancer and other diseases is integrated by the collection and analysis of public data (17). The top 500 genes found in these 3 public databases to be associated with lung adenocarcinoma were gathered for further analysis.

Common genes associated with lung adenocarcinoma

Venn diagram analysis was used to define common genes associated with lung adenocarcinoma from the three public databases. Bioinformatics analysis was subsequently performed on these common genes.

Bioinformatics analysis

Bioinformatics analysis was conducted for all the common genes from the three public databases. GeneMANIA (http://genemania.org/) was used for functional analysis of the interaction network. GeneMANIA can be used to predict the functions of target genes and evaluate the prioritization of genes. Enrichment analysis including Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) were performed to estimate the molecular biological functions of common genes, including biological processes (BPs), cellular components (CCs), and molecular functions (MFs).

Hub gene definition

WebGestalt (http://www.webgestalt.org/) was used to identify the hub genes associated with lung adenocarcinoma from the common genes screened from the three public databases. The WebGestalt database is suitable for seed gene screening with the methods of enrichment analysis and highly significant connected node analysis.

mRNA and protein expression validation

The differences in gene mRNA expression in tumor samples and normal samples were validated using The Cancer Genome Atlas (TCGA) database. Protein expression of hub genes was evaluated by UALCAN (http://ualcan.path.uab.edu/), using both tumor samples and adjacent normal samples of lung adenocarcinoma cases. Immunohistochemistry information was provided by the Human Protein Atlas (HPA, https://www.proteinatlas.org/).

Predictive ability of hub gene for lung adenocarcinoma risk

The prediction ability of hub genes on the risk of lung adenocarcinoma was evaluated by receiver operating characteristic (ROC) curve analysis based on TCGA database. The area under the curve (AUC), sensitivity, and specificity of hub genes were calculated for evaluation.

The correlation between hub genes and immune infiltration

The Tumor-Immune System Interaction Database (TISIDB, http://cis.hku.hk/TISIDB/) was used to estimate the correlation between hub genes and immune infiltration. Furthermore, the association between gene mRNA expression and the abundance of lung adenocarcinoma tumor-infiltrating lymphocytes (TILs) was evaluated. The Pearson correlation coefficient was calculated for evaluation.

Statistical analysis

The mean ± standard deviation (SD) is presented for numerical variables. Student's *t*-test was used to evaluate the differences of means between two groups. P values less than 0.05 (two-sided) was defined as statistically significant. Linear correlation analysis was performed to assess the relationship between gene expression. The SAS software (SAS 9.4) was used for statistical analysis. The R software (R 3.6.1) and GraphPad Prism were used for graphic plotting in this study.

Results

Common genes associated with lung adenocarcinoma screened from public databases

The top 500 genes associated with lung adenocarcinoma from the GeneCards, CTD, and DISEASES databases were identified. Venn diagram analysis identified the 21 common genes associated with lung adenocarcinoma, including *AKT1*, *KRAS*, *EGF*, *ERBB2*, *BRCA1*, *STAT3*, *MMP9*, *EGFR*, *CCND1*, *CDH1*, *CSF3*, *CDKN2A*, *CXCL8*, *MYC*, *NOTCH1*, *GNAS*, *JAK2*, *CTNNB1*, *TP53*, *CD44*, and *VIM* (*Figure 1*).

Bioinformatics analysis

Bioinformatics analysis was conducted for the 21 common genes identified from the three public databases. The results from the STRING database showed an interaction effect among the protein expression of the 21 common genes associated with lung adenocarcinoma. Furthermore, GeneMANIA was performed for functional analysis of the interaction network (*Figure 2* and *Table 1*). The interaction effects of the 21 common genes mainly focused on negative regulation of cell cycle phase transition, the ERBB signaling pathway, regulation of helicase activity, and other important functional aspects.

For enrichment analysis, GO and KEGG analysis were performed for molecular biological functions evaluation

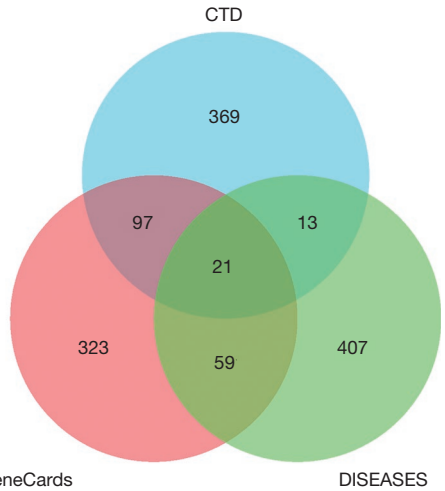

Figure 1 A Venn diagram analysis for the identification of common genes associated with lung adenocarcinoma based on the public databases. CTD, Comparative Toxicogenomics Database.

(*Figure 3* and *Table 2*). For the 21 common genes, the following BPs were enriched: positive regulation of epithelial cell proliferation, glial cell differentiation, and gliogenesis. The following CCs were enriched: apical part of cell, focal adhesion, cell-substrate adherens junction, and cell-substrate junction. The following MFs were enriched: protein phosphatase binding, phosphatase binding, hormone receptor binding, and repressing transcription factor binding.

Hub gene definition

The WebGestalt database was used for hub gene identification based on the 21 common genes associated with lung adenocarcinoma screened from the three public databases. Overrepresentation analysis (ORA) was chosen for the enrichment of hub genes. Genome protein coding

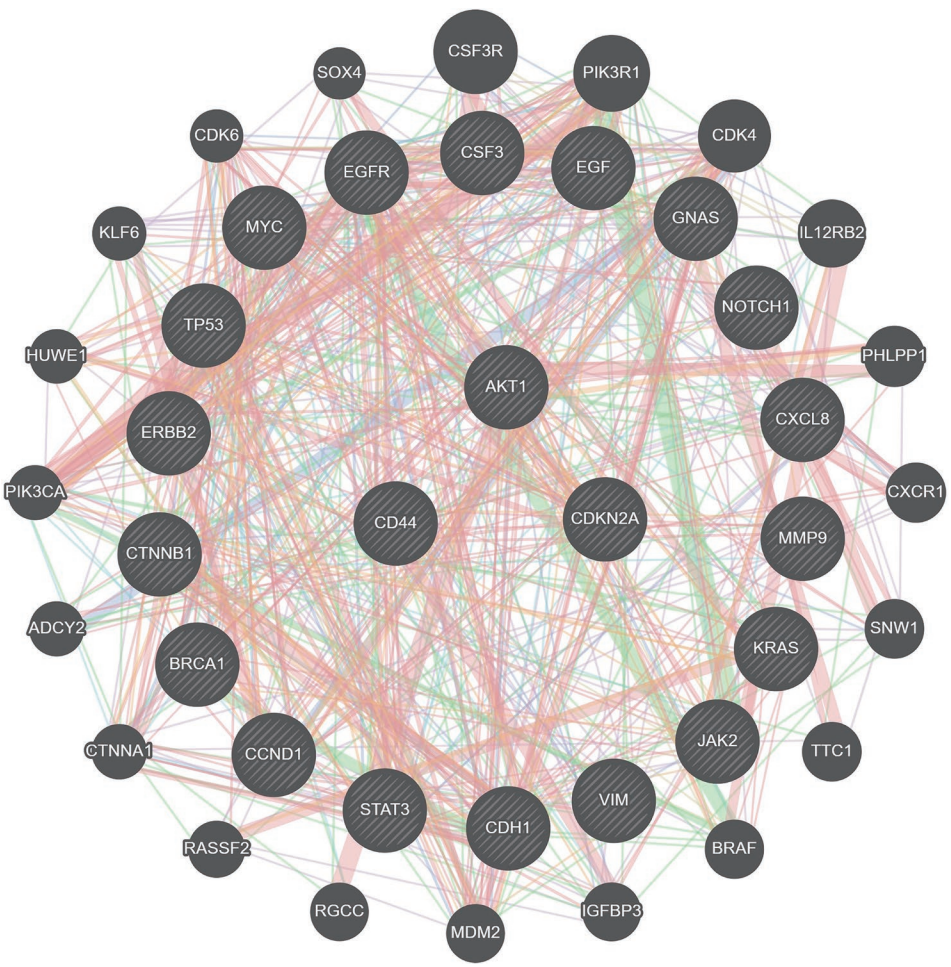

Figure 2 An interaction network analysis built in GeneMANIA using the 21 common genes.

Table 1 The top 10 interaction networks in the 21 common genes according to GeneMANIA

Function	FDR	Genes in network
Regulation of G1/S transition of mitotic cell cycle	1.55E−15	13
Negative regulation of cell cycle G1/S phase transition	4.08E−15	12
Regulation of cell cycle G1/S phase transition	5.91E−15	13
G1/S transition of mitotic cell cycle	1.47E−14	13
Cell cycle G1/S phase transition	1.06E−13	14
Negative regulation of cell cycle phase transition	1.37E−12	13
Negative regulation of mitotic cell cycle	1.13E−11	13
Negative regulation of mitotic cell cycle phase transition	1.74E−11	12
Phosphatidylinositol-mediated signaling	1.94E−10	10
Inositol lipid-mediated signaling	1.67E−09	10

FDR, false discovery rate.

was selected as the reference set. Using a minimum P value criterion for hub gene identification, *AKT1*, *CD44*, and *CDKN2A* were found to have significant interaction effects (*Table 3*).

Hub gene mRNA and protein expression validation

The differences in the expression of the hub genes between lung adenocarcinoma samples and normal sample were evaluated using TCGA database. As shown in *Figure 4*, *CDKN2A* mRNA expression was higher in tumor samples compared to normal samples (1.95±1.50 *vs.* 0.57±0.23, P<0.001). In contrast, the mRNA expression levels of *AKT1* and *CD44* were higher in normal samples compared to tumor samples (3.86±0.51 *vs.* 3.90±0.17, P=0.622 for *AKT1*; and 4.93±0.96 *vs.* 6.04±0.30, P<0.001 for *CD44*). Similar results were observed in 59 tumor samples compared to the corresponding adjacent normal samples (3.85±0.41 *vs.* 3.89±0.17, P=0.498 for *AKT1*; 5.26±0.83 *vs.* 6.03±0.30, P<0.001 for *CD44*; and 2.13±1.42 *vs.* 0.56±0.23, P<0.001 for *CDKN2A*).

The protein expression of hub genes was further assessed

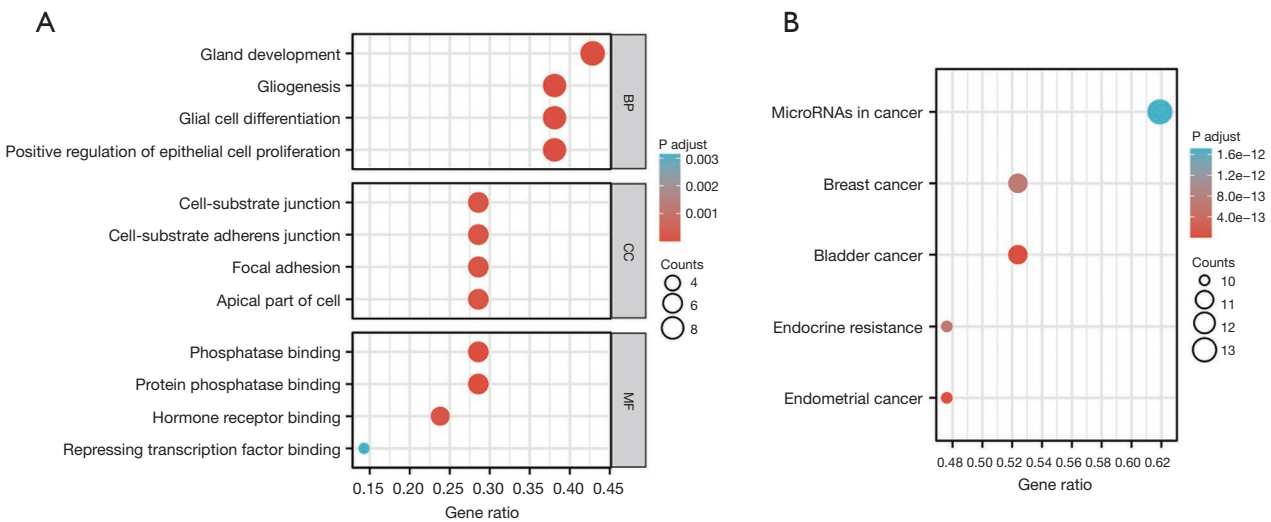

Figure 3 Enrichment analysis of the 21 common genes associated with lung adenocarcinoma. (A) The results of GO analysis. (B) The results of KEGG analysis. BP, biological processes; CC, cellular components; MF, molecular functions; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Table 2 Gene Ontology enrichment analysis of the biological processes, cell components, and molecular functions associated with the 21 common genes in lung adenocarcinoma

Ontology	ID	Description	FDR
BP	GO:0050679	Positive regulation of epithelial cell proliferation	6.21E–08
BP	GO:0010001	Glial cell differentiation	6.21E–08
BP	GO:0048732	Gland development	3.00E–07
BP	GO:0042063	Gliogenesis	3.00E–07
CC	GO:0045177	Apical part of cell	1.19E–04
CC	GO:0005925	Focal adhesion	1.19E–04
CC	GO:0005924	Cell-substrate adherens junction	1.19e–04
CC	GO:0030055	Cell-substrate junction	1.19E–04
MF	GO:0019903	Protein phosphatase binding	1.74E–06
MF	GO:0019902	Phosphatase binding	4.61E–06
MF	GO:0051427	Hormone receptor binding	1.13E–04
MF	GO:0070491	Repressing transcription factor binding	0.003

BP, biological processes; CC, cellular components; MF, molecular functions; FDR, false discovery rate; GO, Gene Ontology.

Table 3 WebGestalt database identification of hub genes among the 21 common genes associated with lung adenocarcinoma

Gene set	Gene set size	Expected value	Enrichment ratio	FDR
PPI_BIOGRID_M255	1,584	1.73	8.07	1.02E–08
PPI_BIOGRID_M501	623	0.68	14.66	8.12E–08
PPI_BIOGRID_M767	144	0.16	38.07	1.98E–06
PPI_BIOGRID_M916	55	0.06	66.44	7.12E–05
PPI_BIOGRID_M747	41	0.04	66.85	0.002
PPI_BIOGRID_M955	10	0.01	182.71	0.007
PPI_BIOGRID_M499	78	0.09	35.14	0.010
PPI_BIOGRID_M923	17	0.02	107.40	0.016
PPI_BIOGRID_M915	25	0.03	73.08	0.032

FDR, false discovery rate.

with UALCAN and HPA. The protein expression levels of AKT1, CD44, and CDKN2A based on UALCAN are presented in *Figure 5*. The protein expression of CDKN2A was higher in tumor samples compared to normal samples (P=0.012), while the protein expression of AKT1 and CD44 was higher in normal samples compared to tumor samples (P=0.014 for AKT1 and P<0.001 for CD44). Immunohistochemical staining of AKT1, CD44 and CDKN2A in the HPA samples showed similar results to that observed with the UALCAN samples (*Figure 6*).

Predictive ability of hub genes for the risk of lung adenocarcinoma

The prediction ability of hub genes for the risk of lung adenocarcinoma was evaluated by ROC curve analysis based on TCGA database (*Figure 7*). The AUC of AKT1, CD44, and CDKN2A for lung adenocarcinoma risk was 0.847 [95% confidence interval (CI): 0.811–0.883], 0.880 (95% CI: 0.850–0.910), and 0.805 (95% CI: 0.770–0.840), respectively, with sensitivity of 89.8%, 93.2%, and 94.9%,

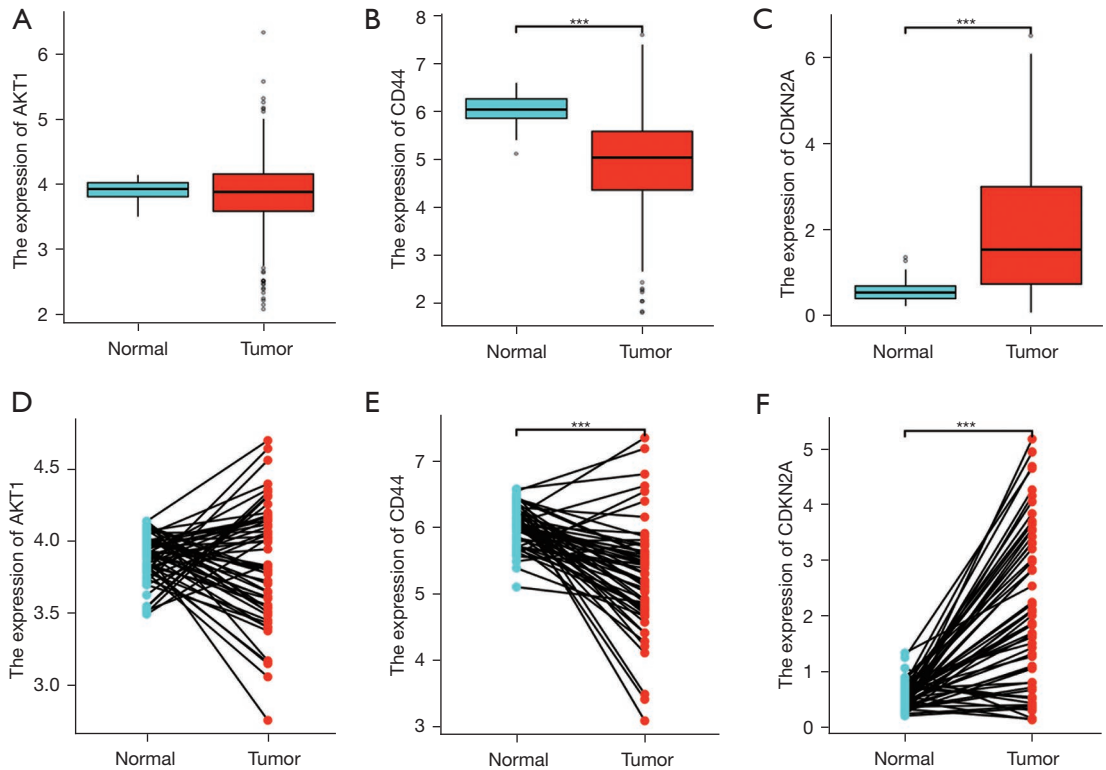

Figure 4 The mRNA expression of hub genes based on TCGA lung adenocarcinoma database. (A-C) Unpaired. (D-F) Paired. ***, P<0.001. TCGA, The Cancer Genome Atlas.

respectively.

The correlation between hub genes and immune infiltration

TISIDB analysis was conducted to estimate the correlation between hub genes and lung adenocarcinoma immune infiltration. *Figure 8* shows that the mRNA expression of *AKT1*, *CD44*, and *CDKN2A* was significantly associated with the abundance of activated (Act-) $CD4^+$ T cells, Act-$CD8^+$ T cells, regulatory T cells (Treg), and CD56 bright cells (P<0.05 for all).

In addition, as shown in *Figure 9*, there were significant differences in the abundance of 4 immune infiltrate cell types, including T cells, effector memory T cells (Tem), Treg cells, and natural killer (NK) CD56 bright cells, depending on high or low expression of *AKT1*, *CD44*, and *CDKN2A*. The results indicated that mRNA expression of *AKT1*, *CD44*, and *CDKN2A* may influence the abundance of TILs in lung adenocarcinoma.

Discussion

Although a large number of bioinformatic databases are available for analysis, the information has not yet been effectively applied to the biological mechanisms, diagnosis, and treatment of disease. Massive datasets with multidimensional information about biology urgently need to be combined and examined from multiple angles (18). In this study, the GeneCards, CTD, and DISEASES databases were used and integrated to explore hub genes correlated with lung adenocarcinoma. According to the results of the Venn Diagram software, 21 genes were defined as common genes among the three databases. Gene enrichment analysis is a common method for functional analysis of a group of genes. Through integration of different databases and further analysis, some crucial biomarkers in disease biological development can be obtained, which is conducive to the exploration of disease pathogenesis, the development of early diagnosis, and the realization of personalized therapy (19,20). Consequently, we combined

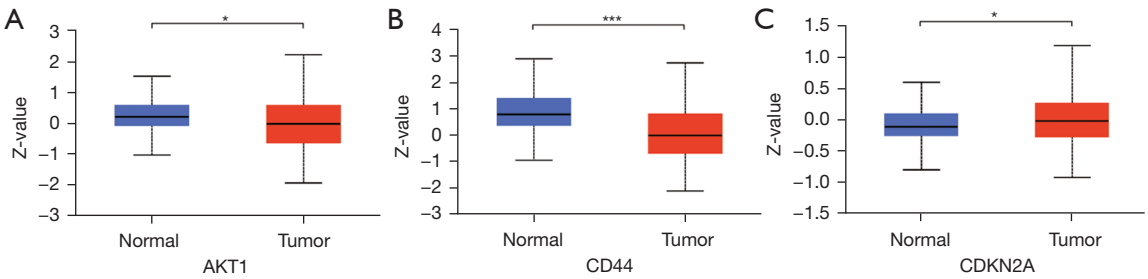

Figure 5 The protein expression of hub genes in lung adenocarcinoma patients based on UALCAN. (A) *AKT1*. (B) *CD44*. (C) *CDKN2A*. *, P<0.05; ***, P<0.001.

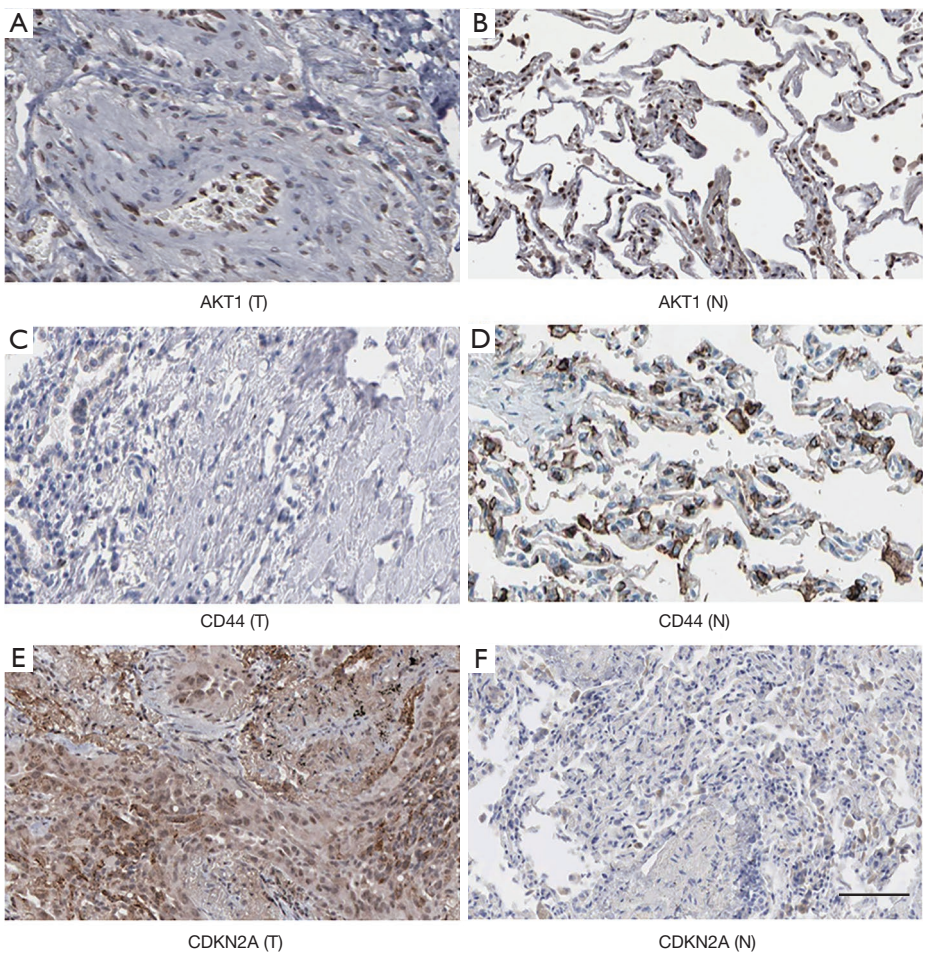

Figure 6 Immunohistochemical staining of hub genes in lung adenocarcinoma patients based on HPA. (A) AKT1, tumor sample (image available from https://www.proteinatlas.org/ENSG00000142208-AKT1/pathology/lung+cancer#img). (B) AKT1, normal sample (image available from https://www.proteinatlas.org/ENSG00000142208-AKT1/tissue/lung#img). (C) CD44, tumor sample (image available from https://www.proteinatlas.org/ENSG00000026508-CD44/pathology/lung+cancer#img). (D) CD44, normal sample (image available from https://www.proteinatlas.org/ENSG00000026508-CD44/tissue/lung#img). (E) CDKN2A, tumor sample (image available from https://www.proteinatlas.org/ENSG00000147889-CDKN2A/pathology/lung+cancer#img). (F) CDKN2A, normal sample (image available from https://www.proteinatlas.org/ENSG00000147889-CDKN2A/tissue/lung#img). The scale bar indicates 100 μm. HPA, Human Protein Atlas; T, tumor sample; N, normal sample.

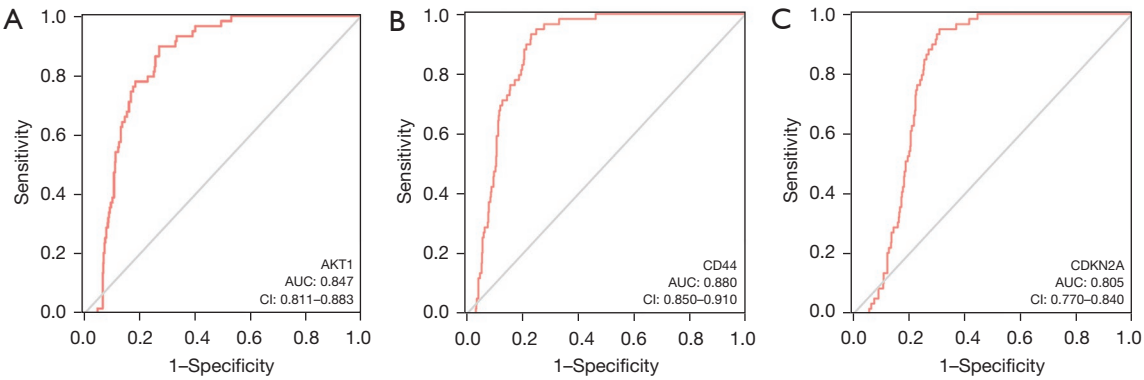

Figure 7 The ROC curve analysis of hub genes for lung adenocarcinoma risk prediction. (A) *AKT1*. (B) *CD44*. (C) *CDKN2A*. AUC, area under the curve; CI, confidence interval; ROC, receiver operating characteristic.

these 21 common genes into a gene set for further bioinformatics analysis. WebGestalt was then used for hub gene identification, and *AKT1*, *CD44*, and *CDKN2A* were identified as having the strongest interaction among the common genes. Furthermore, TCGA, UALCAN, and HPA databases were analyzed for gene expression validation. The results of the ROC curve analysis indicated that *AKT1*, *CD44*, and *CDKN2A* have a good predictive ability on the risk of lung adenocarcinoma. In addition, significant correlations between *AKT1*, *CD44*, and *CDKN2A* and lung adenocarcinoma immune infiltration was observed based on the TISIDB.

As an important member in the AKT kinase family, *AKT1* has been reported to be abnormally expressed in several human tumor cells (21). As a key protein molecule in PI3K/AKT signaling pathway, the protein expression of AKT1 has been proved to be involved in the regulation of various cellular functions, including cell proliferation, metabolism and apoptosis, which is crucial in the development of tumors (22). However, the results have been inconsistent. One study found that *AKT1* can inhibit the migration and invasion of breast cancer cells by regulating the activity of T cells (23). However, another study demonstrated that *AKT1* promoted cell migration and tumor metastasis in a mouse model (24). Rao *et al.* reported that *AKT1* inhibition can promote the migration and invasion of lung adenocarcinoma cells, indicating *AKT1* may have a significant impact on the regulation of tumor metastasis (25). To date, several investigations have suggested that abnormal expression of *AKT1* is significantly associated with clinical features, prognosis, and treatment outcomes in patients with lung adenocarcinoma (26,27). In this current study, *AKT1* was selected as a hub gene associated with lung adenocarcinoma. It was overexpressed in normal tissues compared with tumor tissue. Furthermore, the results from the TISIDB showed that *AKT1* mRNA expression was significantly associated with the abundance of Act-CD4$^+$ T cells and Act-CD8$^+$ T cells, which is consistent with other research results. *AKT1* may be a novel tumor biomarker for the exploration of the biological mechanisms and targeted therapy of lung adenocarcinoma.

Variations in *CD44* expression have been shown to be a major feature in many kinds of human cancers, especially in lung cancer (28,29). However, the conclusions in several studies have been inconsistent. Abnormal expression of CD44 was detected in both blood and tissue samples from lung cancer patients, and was associated with tumor metastasis and prognosis (30). It has also been shown that silencing the gene expression of *CD44* can inhibit cell growth and proliferation in lung cancer, suggesting that *CD44* may have some effect on tumor therapy (31). In addition, *CD44* expression was found to induce resistance to targeted therapy of lung adenocarcinoma through epithelial-mesenchymal transformation (32). Nevertheless, Sung *et al.* reported that *CD44* overexpression can reduce the risk of poor prognosis in lung adenocarcinoma patients (33). Although the biological mechanisms of the influence of *CD44* on lung adenocarcinoma development requires further study, the results in our study suggested that *CD44* may be a potential new tumor biomarker for the evaluation and therapy of lung adenocarcinoma.

CDKN2A (cyclin dependent kinase inhibitor 2A) was first detected in 1994 as a cell cycle regulating factor and was found to be associated with the development of lung cancer (34). Several reports have confirmed that *CDKN2A* expression may be influenced by some crucial genes that

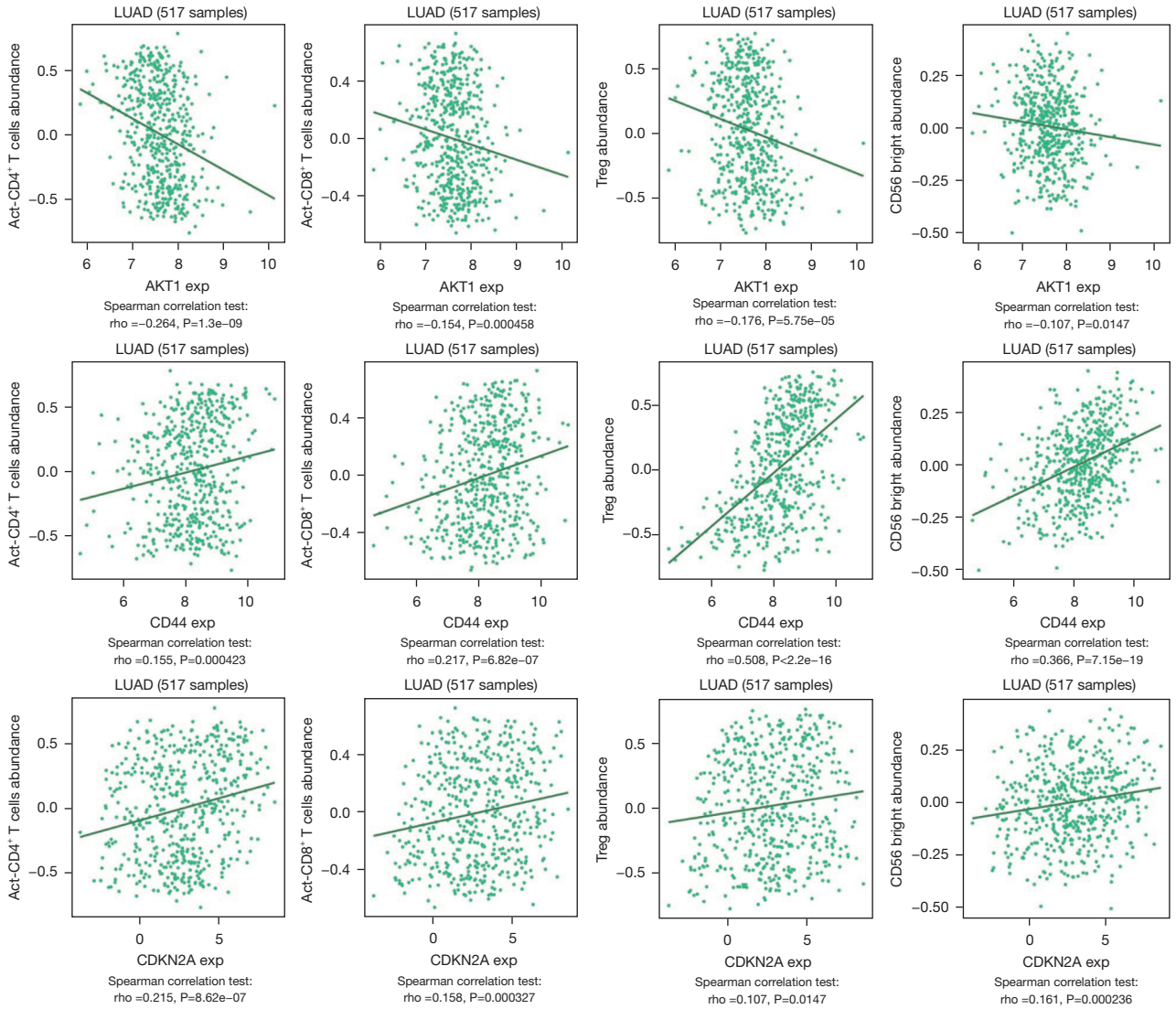

Figure 8 The association between hub gene mRNA expression and lung adenocarcinoma immune infiltration. Act, activated; Treg, regulatory T cells; LUAD, lung adenocarcinoma.

are important in the development of lung adenocarcinoma, such as *MTAP* and *JAK2* (35,36). In addition, mutation of *CDKN2A* has been detected in the early stage of lung cancer, and *CDKN2A* mRNA expression was shown to have good predictive ability for lung cancer risk (37). Gutiontov *et al.* reported that *CDKN2A* mRNA expression can affect the prognosis of lung adenocarcinoma in different clinical features, and has a strong relationship with immunotherapy resistance in lung adenocarcinoma patients (38). Our current study suggested that *CDKN2A* has good predictive effect on lung adenocarcinoma risk. Furthermore, a significant correlation was observed between *CDKN2A* and lung adenocarcinoma immune infiltration. These results suggested that *CDKN2A* may be a potential new biomarker and therapeutic target in patients with lung adenocarcinoma.

Conclusions

In conclusion, using the GeneCards, CTD, and DISEASES databases, a total of 21 genes were identified as common genes associated with lung adenocarcinoma. In addition,

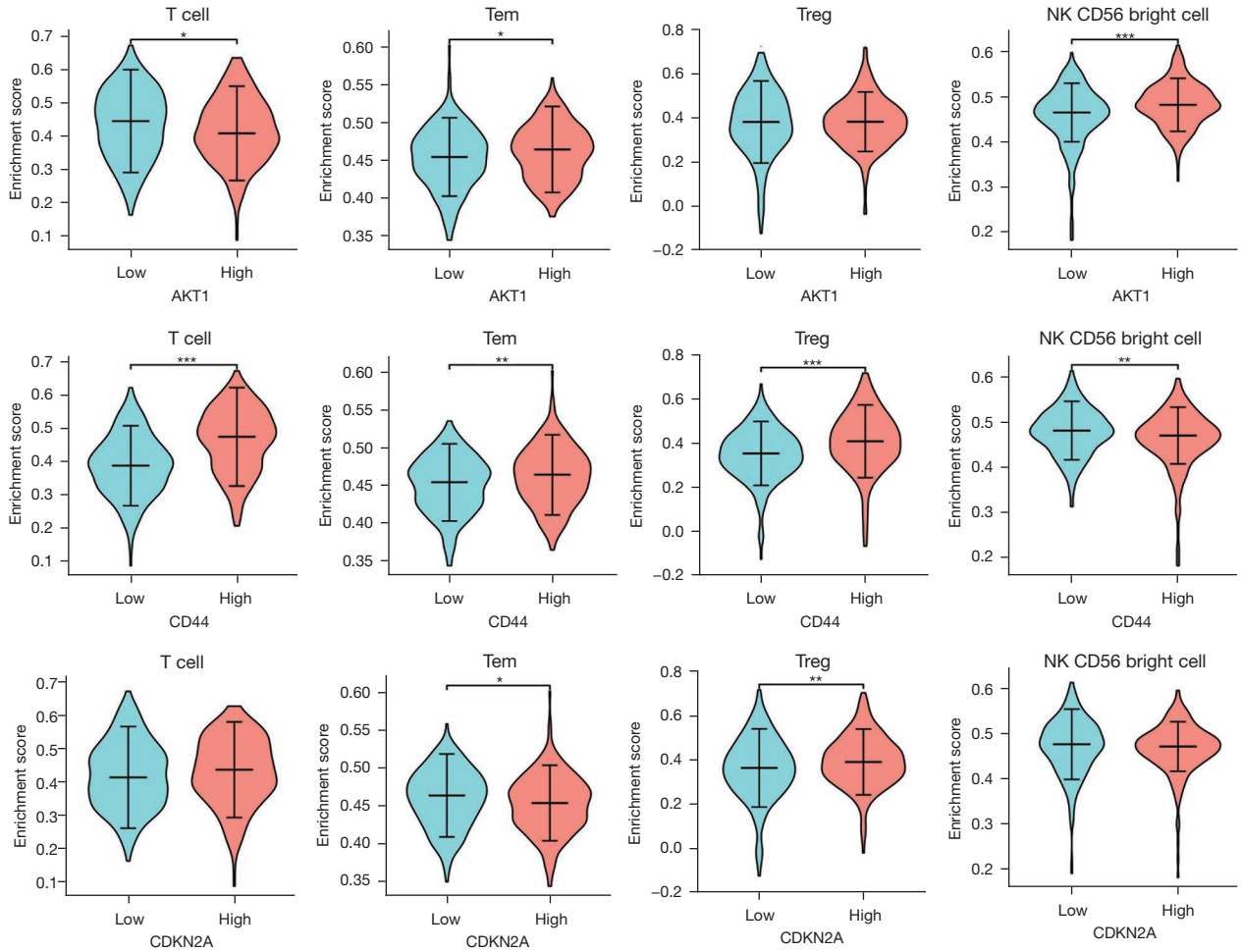

Figure 9 The association between hub gene expression and lung adenocarcinoma immune infiltration in subgroups. *, P<0.05; **, P<0.01; ***, P<0.001. Tem, effector memory T cells; Treg, regulatory T cells; NK, natural killer.

AKT1, *CD44*, and *CDKN2A* were found to be hub genes associated with lung adenocarcinoma. Significant differences in the mRNA and protein expression of hub genes were observed between lung adenocarcinoma samples and normal samples. ROC curve analysis revealed that *AKT1*, *CD44*, and *CDKN2A* have good predictive effect on the risk of lung adenocarcinoma. TISIDB analysis indicated a strong relationship between immune infiltration in lung adenocarcinoma and the expression of *AKT1*, *CD44*, and *CDKN2A*. This investigation identified potential new tumor biomarkers and may provide novel insights into the mechanisms, diagnosis, and individualized treatment of patients with lung adenocarcinoma. The roles of hub genes identified in this study warrant further validation and investigation with regards to the molecular mechanisms of lung adenocarcinoma.

Acknowledgments

The authors are grateful for the availability of data from the public databases.

Funding: This study was supported by the Educational Committee Foundation of Hubei Province (No. Q20212605), the Xiangyang Medical-health Areas Science and Technology Program (No. 2022YL03A), and the Hubei Province Training Program of Innovation and Entrepreneurship for Undergraduates (Nos. S202010510903l and S202210519053).

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at https://tcr.

amegroups.com/article/view/10.21037/tcr-22-2225/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-22-2225/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021. CA Cancer J Clin 2021;71:7-33.
2. Thai AA, Solomon BJ, Sequist LV, et al. Lung cancer. Lancet 2021;398:535-54.
3. Thandra KC, Barsouk A, Saginala K, et al. Epidemiology of lung cancer. Contemp Oncol (Pozn) 2021;25:45-52.
4. Roy-Chowdhuri S. Molecular Pathology of Lung Cancer. Surg Pathol Clin 2021;14:369-77.
5. Travis WD. Lung Cancer Pathology: Current Concepts. Clin Chest Med 2020;41:67-85.
6. Alexander M, Kim SY, Cheng H. Update 2020: Management of Non-Small Cell Lung Cancer. Lung 2020;198:897-907.
7. Sears CR, Mazzone PJ. Biomarkers in Lung Cancer. Clin Chest Med 2020;41:115-27.
8. Mamdani H, Matosevic S, Khalid AB, et al. Immunotherapy in Lung Cancer: Current Landscape and Future Directions. Front Immunol 2022;13:823618.
9. Benusiglio PR, Fallet V, Sanchis-Borja M, et al. Lung cancer is also a hereditary disease. Eur Respir Rev 2021;30:210045.
10. Madama D, Martins R, Pires AS, et al. Metabolomic Profiling in Lung Cancer: A Systematic Review. Metabolites 2021;11:630.
11. Passaro A, Brahmer J, Antonia S, et al. Managing Resistance to Immune Checkpoint Inhibitors in Lung Cancer: Treatment and Novel Strategies. J Clin Oncol 2022;40:598-610.
12. Bodor JN, Boumber Y, Borghaei H. Biomarkers for immune checkpoint inhibition in non-small cell lung cancer (NSCLC). Cancer 2020;126:260-70.
13. Wang Y, Zou S, Zhao Z, et al. New insights into small-cell lung cancer development and therapy. Cell Biol Int 2020;44:1564-76.
14. Wang Y, Zhao Y, Bollas A, et al. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol 2021;39:1348-65.
15. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics 2016;54:1.30.1-1.30.33.
16. Davis AP, Grondin CJ, Johnson RJ, et al. The Comparative Toxicogenomics Database: update 2019. Nucleic Acids Res 2019;47:D948-54.
17. Pletscher-Frankild S, Pallejà A, Tsafou K, et al. DISEASES: text mining and data integration of disease-gene associations. Methods 2015;74:83-9.
18. Zachariou M, Minadakis G, Oulas A, et al. Integrating multi-source information on a single network to detect disease-related clusters of molecular mechanisms. J Proteomics 2018;188:15-29.
19. Boegel S, Castle JC, Kodysh J, et al. Bioinformatic methods for cancer neoantigen prediction. Prog Mol Biol Transl Sci 2019;164:25-60.
20. Richters MM, Xia H, Campbell KM, et al. Best practices for bioinformatic characterization of neoantigens for clinical utility. Genome Med 2019;11:56.
21. Herberts C, Murtha AJ, Fu S, et al. Activating AKT1 and PIK3CA Mutations in Metastatic Castration-Resistant Prostate Cancer. Eur Urol 2020;78:834-44.
22. Jiang N, Dai Q, Su X, et al. Role of PI3K/AKT pathway in cancer: the framework of malignant behavior. Mol Biol Rep 2020;47:4587-629.
23. Yoeli-Lerner M, Yiu GK, Rabinovitz I, et al. Akt blocks breast cancer cell motility and invasion through the transcription factor NFAT. Mol Cell 2005;20:539-50.
24. Cho JH, Robinson JP, Arave RA, et al. AKT1 Activation Promotes Development of Melanoma Metastases. Cell Rep 2015;13:898-905.
25. Rao G, Pierobon M, Kim IK, et al. Inhibition of AKT1

signaling promotes invasion and metastasis of non-small cell lung cancer cells with K-RAS or EGFR mutations. Sci Rep 2017;7:7066.
26. Zhang G, Chen HX, Yang SN, et al. MAGI1-IT1 stimulates proliferation in non-small cell lung cancer by upregulating AKT1 as a ceRNA. Eur Rev Med Pharmacol Sci 2020;24:691-8.
27. Malanga D, De Marco C, Guerriero I, et al. The Akt1/IL-6/STAT3 pathway regulates growth of lung tumor initiating cells. Oncotarget 2015;6:42667-86.
28. Jiang YX, Siu MK, Wang JJ, et al. Ascites-derived ALDH+CD44+ tumour cell subsets endow stemness, metastasis and metabolic switch via PDK4-mediated STAT3/AKT/NF-κB/IL-8 signalling in ovarian cancer. Br J Cancer 2020;123:275-87.
29. Kong T, Ahn R, Yang K, et al. CD44 Promotes PD-L1 Expression and Its Tumor-Intrinsic Function in Breast and Lung Cancers. Cancer Res 2020;80:444-57.
30. Wang Y, Guo Y, Lin H, et al. Expression of CD44 in Tumor Tissue and Serum of Small Cell Lung Cancer and Its Clinical Prognostic Significance. Zhongguo Fei Ai Za Zhi 2021;24:583-90.
31. Wang YY, Vadhan A, Chen PH, et al. CD44 Promotes Lung Cancer Cell Metastasis through ERK-ZEB1 Signaling. Cancers (Basel) 2021;13:4057.
32. Suda K, Murakami I, Yu H, et al. CD44 Facilitates Epithelial-to-Mesenchymal Transition Phenotypic Change at Acquisition of Resistance to EGFR Kinase Inhibitors in Lung Cancer. Mol Cancer Ther 2018;17:2257-65.
33. Sung WJ, Park KS, Kwak SG, et al. Epithelial-mesenchymal transition in patients of pulmonary adenocarcinoma: correlation with cancer stem cell markers and prognosis. Int J Clin Exp Pathol 2015;8:8997-9009.
34. Nobori T, Miura K, Wu DJ, et al. Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. Nature 1994;368:753-6.
35. Liu W, Zhuang C, Huang T, et al. Loss of CDKN2A at chromosome 9 has a poor clinical prognosis and promotes lung cancer progression. Mol Genet Genomic Med 2020;8:e1521.
36. Horn S, Leonardelli S, Sucker A, et al. Tumor CDKN2A-Associated JAK2 Loss and Susceptibility to Immunotherapy Resistance. J Natl Cancer Inst 2018;110:677-81.
37. Wang P, Wang F, He H, et al. TP53 and CDKN2A mutations in patients with early-stage lung squamous cell carcinoma: an analysis of the correlations and prognostic outcomes. Ann Transl Med 2021;9:1330.
38. Gutiontov SI, Turchan WT, Spurr LF, et al. CDKN2A loss-of-function predicts immunotherapy resistance in non-small cell lung cancer. Sci Rep 2021;11:20059.

Cite this article as: Zeng C, Zhou Y, Ye W, Fang Z, Wang K. Exploration and validation of hub genes in lung adenocarcinoma based on bioinformatics analysis. Transl Cancer Res 2022;11(10):3814-3826. doi: 10.21037/tcr-22-2225

