Proceedings of ICMMGH 2026 Symposium: Cell and Molecular Biology: Understanding Immune Cell Functions
DOI: 10.54254/2753-8818/2026.PJ30522

The Application of Bioinformatics in the Research of Lung Adenocarcinoma

Licheng Lu

School of Biotechnology, East China University of Science and Technology, Shanghai, China
24012320@mail.ecust.edu.cn

Abstract. Lung adenocarcinoma is a major subtype of non-small cell lung cancer, and both morbidity and mortality are very high in China. Lung adenocarcinoma has significant tumor heterogeneity and treatment resistance, which poses a severe challenge to clinical diagnosis and treatment. Due to the significant differences in etiology and prognosis among patients, traditional treatment methods such as surgical resection, radiotherapy, and chemotherapy have limited effects and significant side effects. The rapid development of bioinformatics in recent years has made up for this deficiency in the treatment of lung adenocarcinoma. By integrating multi-omics data, bioinformatics not only deepens the understanding of the pathogenesis of lung adenocarcinoma but also provides new perspectives for the optimization of clinical diagnosis and treatment strategies. This paper systematically expounds the application of bioinformatics in molecular typing, key gene identification, therapeutic target identification and prognosis assessment of lung adenocarcinoma, and summarizes the future development directions, in order to provide theoretical basis related to the diagnosis of this disease.

Keywords: Biomarkers. Bioinformatics, Lung adenocarcinoma, Multiomics, Precision medicine

1. Introduction

Lung cancer, one of the most common and fatal malignancies worldwide, is mainly divided into non-small cell lung cancer and small cell lung cancer, with non-small cell lung cancer accounting for 85%. It can be further classified into adenocarcinoma, squamous cell carcinoma, large cell carcinoma, etc. Lung adenocarcinoma is the most common type of non-small cell lung cancer, accounting for approximately 40% to 50% of all lung cancer cases [1]. Lung adenocarcinoma is highly heterogeneous and its pathogenesis is complex, involving multiple gene mutations and abnormal signaling pathways, making it difficult for traditional diagnosis and treatment models to achieve individualized and precise treatment.

With the rapid development and cost reduction of high-throughput sequencing technology, cancer research has entered the big data era. A single sample can generate terabytes of genomic, transcriptomic, epigenomic and proteomic data [2]. How to extract valuable biological information from these massive amounts of data has become a key issue in current lung adenocarcinoma research. Bioinformatics, as an interdisciplinary subject that combines biology, mathematics, and

DOI: 10.54254/2753-8818/2026.PJ30522

computer science, provides powerful tools for parsing multi-omics data by developing and applying computational methods.

The application of bioinformatics in lung adenocarcinoma research is mainly reflected in three aspects: First, the identification of key genes and signaling pathways through methods such as differential expression analysis and functional enrichment analysis; Second, using machine learning algorithms to construct prognostic prediction models to achieve risk stratification; Third, by integrating multi-omics data to reveal drug response mechanisms and achieve individualized treatment [3]. These applications have not only accelerated the progress of basic research on lung adenocarcinoma but also facilitated the implementation of the concept of precision medicine in clinical practice.

This article will systematically elaborate on the application of bioinformatics in key gene identification, molecular typing, prognostic model construction and treatment strategy exploration of lung adenocarcinoma, and look forward to future development trends, with the aim of providing new ideas for the precise prevention and treatment of lung adenocarcinoma.

2. Key applications of bioinformatics in lung adenocarcinoma research

Cluster analysis based on bioinformatics divides lung adenocarcinoma into different molecular subtypes, providing a scientific basis for precise diagnosis and treatment.

2.1. Molecular typing and precise judgment

2.1.1. Typing based on transcriptome and proteome

Traditional gene expression profiling, such as the American Cancer Genome Atlas Project, classifies lung adenocarcinoma into subtypes such as proximal proliferative, proximal inflammatory, and terminal respiratory unit, each with unique clinicopathological characteristics and prognosis. Patients with proximal proliferative lung adenocarcinoma have a lower five-year survival rate, while those with terminal respiratory unit type have a better prognosis. In recent years, the integration of proteomics has further refined the classification. For example, a proteomic study of KRAS G12-mutated lung adenocarcinoma, through bioinformatics clustering, revealed three molecular subtypes with different types of point mutations in the amino acid at codon 12, one of which was a KRAS G12C mutation with a population concentration tendency, providing a key entry point for precision treatment [4].

2.1.2. Immunotyping based on spatial characteristics of the tumor microenvironment

The tumor microenvironment is like a complex ecosystem whose spatial structure has a profound impact on disease progression and treatment response. A breakthrough study published in Nature Communications used an AI-driven spatial cell phenomics approach to systematically analyze the tumor microenvironment of 1,168 patients with non-small cell lung cancer. The method analyzed complex spatial relationships of 43 cell phenotypes simultaneously using multiplex immunofluorescence imaging and deep learning models. The study identified 10 different cellular niches in lung adenocarcinoma, such as the hot niches rich in lymphocytes and the cold niches (immune-exempt) with almost no immune cell infiltration. Based on this spatial typing model, researchers improved the accuracy of lung adenocarcinoma risk stratification by 14% compared to traditional TNM staging [5]. More importantly, the method was able to identify high-risk subgroups

DOI: 10.54254/2753-8818/2026.PJ30522

with poor prognosis and potential to benefit from adjuvant chemotherapy in patients traditionally staged at stage 1, showing great potential for clinical translation.

2.2. Discovery of key genes and molecular markers

Bioinformatics methods systematically reveal key molecular events in the occurrence and development of lung adenocarcinoma by integrating multi-omics data.

2.2.1. Driver gene mutation lineages and population heterogeneity

Extensive data analysis based on databases such as TCGA and GEO confirmed significant racial, gender, and smoking status differences in mutations of key driver genes for lung adenocarcinoma. In the Asian population of lung adenocarcinoma, women, non-smokers, and patients with advanced clinical stage have a higher frequency of EGFR mutations. And the racial heterogeneity of this mutation rate is very obvious, with a much higher EGFR mutation rate in the Asian population than in the European population [6].

2.2.2. Screening of prognostic pivot genes

Through standard bioinformatics processes (differential expression analysis, PPI network construction, survival analysis), researchers were able to screen out hubgenes closely related to prognosis from massive data. For example, one study screened out hub genes including GNG11 and CXCR2 by analyzing three GEO datasets and confirmed that high expression of most of these genes was significantly associated with overall survival in patients with lung adenocarcinoma [7]. Another study found that genes such as ASPM, CEP55, and TOP2A were significantly highly expressed in lung adenocarcinoma tissues, and their high expression was closely associated with poor prognosis in patients. These genes are mainly involved in biological processes such as cholesterol metabolism, providing important references for the study of the mechanism of lung adenocarcinoma [8].

2.2.3. In-depth analysis of immunotherapy markers

Bioinformatics analysis revealed complex associations between PD-L1 expression levels, tumor mutational burden (TMB), and immune cell infiltration characteristics in the tumor microenvironment and immunotherapy responses. For example, high infiltration of CD8+ T cells is often associated with a better response to immunotherapy. Notably, the aforementioned study on DUSP22 also found that the absence of DUSP22 simultaneously leads to increased PD-L1 expression and activation of the c-Met receptor, which explains why clinical trials combining EGFR and PD-L1 inhibitors failed to produce synergistic effects [9]. The treatment strategy for lung adenocarcinoma needs to take into account the intermolecular regulatory network.

2.3. Therapeutic target identification and drug response prediction

Use title case for the paper title. Function words like “at”, “or”, “with”, etc. should not be capitalized unless they are the first word of the title.

Bioinformatics, through multi-omics data integration and computational modeling, has greatly advanced the precision treatment of lung adenocarcinoma.The application of bioinformatics in the treatment of lung adenocarcinoma is mainly reflected in the discovery of novel therapeutic targets and the prediction of drug responses. By integrating multi-omics data, researchers identified

DOI: 10.54254/2753-8818/2026.PJ30522

multiple potential therapeutic targets, such as abnormal activation of genes like MET, RET, NTRK, etc., providing new directions for targeted drug development [4]. Computational models based on gene expression characteristics were able to predict the sensitivity of tumor cells to specific drugs. For example, ERCC1 expression levels are associated with platinum-based drug sensitivity. Studies have found that tislelizumab injection combined with anlotinib hydrochloride capsules can effectively inhibit the expression of the drug resistance gene ERCC1, which provides an idea for improving the clinical efficacy of chemotherapy drugs [10]. Bioinformatics analysis of samples before and after treatment revealed mechanisms related to drug resistance in organisms. Studies have shown that the activation of the epithelial-mesenchymal transition pathway and alterations in the tumor microenvironment are also important mechanisms of resistance [11]. These findings offer new therapeutic strategies for overcoming drug resistance.

2.4. Construction and validation of prognostic models

Bioinformatics builds increasingly accurate prognostic models for lung adenocarcinoma by integrating multi-dimensional data.

2.4.1. Prognostic models based on specific gene sets

Screening key genes associated with the prognosis of lung adenocarcinoma through bioinformatics methods and constructing risk scoring models based on this is one of the current mainstream research strategies. Chen Shaoming et al. focused on pyroptosis, a particular type of programmed cell death, and obtained lung adenocarcinoma gene expression data from public databases such as TCGA, using methods such as differential expression analysis, Cox regression, and LASSO regression A prognostic model consisting of four pyroptosia-related genes (CPA3, FAT1, MST1, and TFAP2A) was successfully constructed. The study divided lung adenocarcinoma patients into high-risk and low-risk groups based on median risk values, and survival analysis showed that patients in the high-risk group had a significantly poorer prognosis. Further evaluated by receiver operating characteristic curves, the study confirmed that the model had a good predictive ability for patient survival. The model demonstrated robust predictive performance on the training set, validation set, and multiple external datasets including GSE30219, GSE31210, and GSE50081, indicating its reliability [12].

2.4.2. Applications of nomogram fused with clinicopathological features

To enhance the clinical practicality and prediction accuracy of prognostic models,combining genetic characteristics with clinicopathological information to construct visual nomographs is another important development direction. A study published in the Journal of Mathematical Medicine obtained clinicopathological information and immune scores of 337 patients with lung adenocarcinoma from the TCGA database. The researchers first divided the patients into three subgroups with different prognostic characteristics based on their immune scores and found that patients with high immune scores had significantly better overall survival than those with low scores. Subsequently, through multivariate COX proportional hazards regression model analysis, the researchers identified immune scores and tumor TNM stage as independent predictors of overall survival in patients with lung adenocarcinoma. Based on these two key metrics, the researchers constructed a nomogram model that predicted a C-index of 0.71 for overall survival. By plotting calibration curves for 1-year, 3-year, and 5-year overall survival probabilities, it was found that there

DOI: 10.54254/2753-8818/2026.PJ30522

was a significant consistency between the predicted values of the nomogram and the actual observations of the patients, indicating that the model had good predictive accuracy [13].

3. Ensemble learning models based on machine learning algorithms

With the development of artificial intelligence technology, using advanced machine learning algorithms to mine deep information in gene expression data is an effective approach for predicting the prognosis of lung adenocarcinoma. One study presents an ensemble learning approach based on Light GBM (Light Gradient Boosting Machine). The study first downloaded RNA-seq data and clinical data of lung adenocarcinoma from the TCGA database and conducted differential expression analysis using the Limma package to screen out differentially expressed immune genes associated with lung adenocarcinoma. Then, using methods such as Cox proportional hazards model, XGBoost algorithm and LASSO regression, 17 immune-related genes were finally screened out for the construction of the prognostic model. The researchers used the K-means algorithm to divide patients into three subgroups with different survival risks, with 5-year survival rates above 65%, between 30% and 65%, and below 30%, respectively. And the area under the receiver operating characteristic curve of the output results of the LightGBM model was as high as 96%, 98%, and 96%, indicating that the model can accurately stratify the risk of lung adenocarcinoma patients and predict their five-year survival rate [14].

4. Conclusion

Bioinformatics has achieved remarkable results in lung adenocarcinoma research,mainly in four aspects: First, the molecular typing system of lung adenocarcinoma based on molecular characteristics has promoted the development of precision diagnosis; Second, through the integration and analysis of multi-omics data, a large number of key genes and molecular markers related to the occurrence and development of lung adenocarcinoma were discovered; Third, in terms of treatment, it provides an important basis for the discovery of new targets and the prediction of drug responses; Fourth, an accurate prognostic prediction model was constructed using machine algorithms.

However, there are still some challenges in the field: first of all, most studies are based on retrospective data and need to be verified by prospective studies; Secondly, the integration and standardization of data from different platforms still need to be improved; In addition, the efficiency of clinical translation of bioinformatics findings needs to be improved, and data privacy and bioethics are also potential problems.

The future development directions of bioinformatics in lung adenocarcinoma research include: Advancing the deep integration of multi-omics data, combined with new technologies such as single-cell sequencing and spatial transcriptomics, to analyze lung adenocarcinoma heterogeneity at higher resolution; Develop more advanced artificial intelligence algorithms to improve the accuracy of molecular typing and prognosis prediction; Establish standardized bioinformatics analysis processes to facilitate the clinical translation of research results; Strengthen multi-center collaboration and build a larger multi-omics database of lung adenocarcinoma; Focus on the development of bioinformatics talents and promote interdisciplinary integration.

With the continuous advancement of technology and the improvement of research methods, bioinformatics is bound to play a more important role in the precise prevention and treatment of lung adenocarcinoma. Through continuous technological innovation and multidisciplinary

DOI: 10.54254/2753-8818/2026.PJ30522

collaboration, bioinformatics is expected to provide more individualized treatment plans for patients with lung adenocarcinoma and ultimately improve their prognosis.

References

[1] Qian Xiaoyu. A Study on the Predictive Value of PD-L1 expression and LIPI Score in Immunotherapy Combined with chemotherapy for driver Gene-negative Advanced Lung Adenocarcinoma [D]. Inner Mongolia medical university, 2024. DOI: 10.27231 /, dcnki. Gnmyc. 2024.000171.

[2] Stephens Z D, Lee S Y, Faghri F, et al. Big data: astronomical genomical? [J].PLoS Biology, 2015, 13(7): e1002195.

[3] Peng Huiyi, Wang Xuefeng, Deng Biao, et al. Analysis of the relationship between basement membrane immune characteristic genes and lung adenocarcinoma based on machine learning and bioinformatics [J]. Journal of Qiqihar Medical College, 25, 46(19): 1801-1812.

[4] Shi X, Hu L, Huang Y, Zhang MF, Ying X, Yuan X, Wen N, Lu J, Zou H, Tan X, He QY, Wang F, Yang H, Zhang CZ. Integrative proteomic characterization of human lung adenocarcinoma with KRAS G12mutations reveals molecular Pathogenesis. J Adv Res. 2025 Sep 14: S2090-1232 (25), 00697-6.Doi: 10.1016 / j.j. Motorcycle 2025.09.014. Epub ahead of print. PMID: 40957478.

[5] Schallenberg S, Dernbach G, Ruane S, Jurmeister P, Bohm C, Standvoss K, Ghosh S, FrentschM, Dragomir MP, Keyl PG Friedrich C, Na IK, Merkelbach-Bruse S, Quaas A, Frost N, Boschung K, Randerath W, Schlachtenberger G, Heldwein M, Keilholz U, Hekmat K, Ruckert JC, Buttner R, Vasaturo A, Horst D, Ruff L, Alber M, Muller KR Klauschen F. AI-powered spatial cell phenomics enhances risk stratification in non-small cell lung cancer. Nat Commun. 2025 Nov 3; 16(1): 9701.doi: 10.1038/s41467-025-65783-z. PMID: 41184299; PMCID: PMC12583542.

[6] Yang Xueqi, Du Zhenwu, Liu Bohong, et al. A Study on the Relationship between gene mutations and clinicopathological characteristics in patients with non-small cell lung cancer based on high-throughput sequencing [J]. International Journal of Laboratory Medicine, 25, 46(21): 2613-2620.

[7] Liu Xiaodong, Sun Zhanbing, Jin Huiyun, et al. Based on bioinformatics screening key genes related to prognosis of lung adenocarcinoma [J]. Journal of central southmedical science magazine, 2025 does (01) : 12-16. DOI: 10.15972 / j.carol carroll nki. 43-1509 / r. 2025.01.003.

[8] Li Jingru, Cao Bangwei. Cholesterol metabolism in patients with lungadenocarcinoma subtypes of exploration and survival prediction model building [J].Journal of the China hospital drug use evaluation and analysis, 2024, 24 (12) : 1409-1412. The DOI: 10.14009 / j.i SSN. 1672-2124.2024.12.001.et al. DUSP22 inhibits lung tumorigenesis by

[9] Lin, HH., Chang, CW., Liao YT. suppression of EGFR/c-Met signaling. Cell Death Discov. 10, 285 (2024).https: //doi.org/10.1038/s41420-024-02038-8

[10] Wang Huanhuan, Zhang Xiaohang, Wang Xiyong. Clinical study of tislelizumab combined with anlotinib in patients with stage III-IV lung adenocarcinoma [J].Chinese journal of clinical pharmacology, 2025, 9 (12) : 1651-1657. The DOI: 10.13699/ j.carol carroll nki. 1001-6821.2025.12.001.

[11] Shen Zhiyong, Lin Xiuhua, Wang Ling, et al. FEN1 knockdown inhibits tumor growth and its molecular mechanism in Lewis lung cancer mouse models [J]. Journal of Clinical Pulmonology, 2020, 30(11): 1706-1713.

[12] Chen Shaoming, Hu Yan, Hong Xudong, et al. Based on machine learning model construction and prognosis of lung adenocarcinoma cancer cell jiao wu [J]. Journal of the Chinese general medicine, 2025, 23 (01) : 21-25. DOI: 10.16766 / j.carol carroll nki. Issn 1674-4152.003827.

[13] Yao Wei, Wang Pingfei. Construction of prognostic prediction nomomaps for lung adenocarcinoma patients based on immune scores [J]. Chinese Journal of Mathematics and Medicine, 24, 37(09): 649-656.

[14] Meng X, Tian Y Zhang X. [Screening of immune related gene and survival prediction of lung adenocarcinoma patients based on LightGBM model]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi. 2024 Feb 25; 41(1): 70-79. Chinese.doi: 10.7507/1001-5515.202305038.PMID: 38403606; PMCID: PMC10894725.