International Journal of General Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Quantitative Analysis of TP53-Related Lung Cancer Based on Radiomics

Hongyu Qiao[1], Zhongxiang Ding[2], Youcai Zhu[1], Yuguo Wei[3], Baochen Xiao[1], Yongzhen Zhao[1], Qi Feng[2]

[1]Zhejiang Rongjun Hospital, Jiaxing, People's Republic of China; [2]Department of Radiology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, People's Republic of China; [3]GE Healthcare Life Sciences, Hangzhou, People's Republic of China

Correspondence: Qi Feng, Tel +86-13588764520, Email fengqi0707@126.com

Background: The role of TP53 mutations in the diagnosis and treatment of lung cancer has attracted increasing attention from experts worldwide. This study aimed to explore the expression of *TP53* gene in lung cancer and its correlation with radiomics quantitative features.

Methods: A total of 93 cases of lung cancer confirmed by pathology were selected, including 44 cases with TP53 mutations and 49 cases with TP53 wild-type. ITK-SNAP software was used to segment the pulmonary nodules, AK software was used to extract radiomic features, and a model was established to predict the type of *TP53* gene mutation in lung cancer lesions.

Results: A total of 852 features were extracted, and 10 features remained after feature selection. The accuracy, areas under the curve, specificity, sensitivity, positive predictive value, and negative predictive value of the logistic regression model were 0.80, 0.86, 0.89, 0.74, 0.90, and 0.71, respectively.

Conclusion: *TP53* gene mutations are correlated with radiomic features in lung cancer, which may have application value for TP53 therapy in the future.

Keywords: radiomics, TP53, lung cancer, mutation

Introduction

On September 12, 2018, WHO published the "Global Cancer Report 2018". Lung cancer still ranked first in the incidence and cause of death of cancer.[1] In 2018, there were more than 2 million lung cancer patients worldwide and 1.76 million deaths due to lung cancer.[1] About 70% of the lung cancer patients are at an advanced stage when diagnosed, and the 5-year survival rate is less than 16%.[2]

Up to 50% of the human cancers contain mutations on both TP53 alleles, and TP53 mutations are found in 96% of the ovarian cancers,[3] 54% of the invasive breast cancers,[4] 86% of the small-cell lung cancers,[5] and 75% of the pancreatic cancers.[6] TP53 mutations confer a predisposition to the early onset of a variety of mesenchymal and epithelial tumors. *TP53* gene is a tumor suppressor gene encoding tumor protein p53, and *TP53* gene is also a well-known candidate gene for predicting chemotherapy response. TP53 mutation is more common in non-small cell lung cancer (NSCLC) patients, and its targeted therapy research has attracted the attention of many experts worldwide. Since TP53 mutation is present in nearly half of NSCLC and the incidence of NSCLC is increasing year by year, eliminating its oncogenic effect is undoubtedly of great significance for the successful treatment of NSCLC. The research on *TP53* gene targeted therapy has attracted the attention of global experts. And making good progress. For example, recent studies have found that the Gof mutant p53 can be used as a therapeutic target.[7]

Radiomics is based on conventional imaging diagnosis and deep mining of data information to find out the subtle structural features of diseases, so as to reflect the changes in human tissues, cells and genes, thereby providing help for the diagnosis and treatment of diseases.[8] Radiomic feature may provide superior diagnostic performance compared with those of morphologic CT features and radiologists.[9] Radiomics is considered to be a promising quantitative tool for the characterization of lung lesion phenotypes and uses large amounts of quantitative CT image features.[10] If radiomic

Received: 26 October 2022
Accepted: 25 November 2022
Published: 6 December 2022

features or conventional signs can be used to identify NSCLC patients with TP53 mutation, it will be of great value to guide clinical practice. The aim of this study is to explore the expression of TP53 in lung cancer and its correlation with radiomics quantitative features.

Materials and Methods

Patients

A retrospective method was used to select 93 patients with lung cancer confirmed by pathology in Zhejiang Rongjun Hospital from December 2017 to December 2021, including 44 cases of TP53 mutation and 49 cases of TP53 wild type by gene detection. Genetic testing uses second-generation sequencing (next-generation sequencing (NGS)). All specimens were derived from tissue samples. There were 30 males and 14 females in the mutation group, aged from 53 to 87 years, with an average of 70.5±10.1 years. There were 26 males and 23 females within the wild-type group, aged from 42 to 87 years, with an average of 66.7±13.6 years. All patients signed a written informed consent form before the examination. This retrospective study involving human participants was reviewed and approved by Zhejiang Rongjun Hospital Ethics Committee (No. 2022–071). Our study was conducted in accordance with the Declaration of Helsinki.

Lung CT Examination

Philips incisive CT was used for lung CT scanning, using lung scanning mode, ranging from the apex to the bottom of the lung, slice thickness of 5mm, slice spacing of 5mm, reconstructed slice thickness of 2.5mm, pitch 1.2, matrix 512×512, tube voltage 120kv, automatic ma/s technology, iterative algorithm image reconstruction. The view of field (FOV) was 350mm. The pulmonary window width was 1500 and the window level was −400.

Radiomic Feature Extraction from Lung CT

The flow chart of this study is shown in Figure 1. ITK-SNAP software was used for image segmentation. The lung CT image with slice thickness of 2.5mm was selected. The region of interest (ROI) was placed layer by layer along the contour of the lung nodules in all layers including the selected lung nodules, and the ROI was further adjusted by removing the calcification area. The original images and segmented three-dimensional ROI images of pulmonary nodules were imported into the Artificial Intelligent Kit (AK) software. Image preprocessing was performed using AK software to adjust the image resolution to 1mm×1mm×1mm for resampling. The image gray level was then uniformly adjusted to 0–255 for normalization. AK platform was used in calculating characteristics. In this project, the features to be calculated include histogram features, Gray Level Co-occurrence Matrix (GLCM), Gray Level Dependence Matrix (GLDM), and Neigbouring Gray Tone Difference Matrix (NGTDM), Gray Level Run Length Matrix (GLRLM), Gauess Transform and Wavelet Features. Data were processed by AK software for outlier processing, data partitioning, and data preprocessing. The histogram features are mainly based on the statistics of gray or brightness information of the lesion area, and then the overall distribution of gray information of the lesion area is studied. Shape parameters mainly describe the shape characteristics of the lesion area mathematically and attempt to describe the shape compactness of the lesion. GLCM obtains the co-occurrence matrix based on statistical pixels with different directions and probabilities, and then quantifies the distribution of the co-occurrence matrix to describe the complexity, change degree, texture thickness and other information of the lesions.[11] Haralick's feature is also based on GLCM to extract the corresponding features, selecting a GLCM with four directions (0°, 45°, 90°, 135°) with a step size of 1, and then calculating the sum of the average values.[12] The characteristic of RLM is to obtain the length matrix by calculating the probability of pixels that appear continuously in different directions and step sizes to describe the complexity, degree of variation and texture thickness of the lesion.[13] After calculation, a total of 852 features were extracted. Labels "0" and "1" indicating mutant and wild groups were added to each subject data.

Feature Selection

Feature selection was performed twice for each of the two groups of nodules based on the calculated features. The preprocessing was divided into three steps: outliers were replaced by mean values; the proportion of training set was 0.7, and the proportion of test set was 0.3; and unit restrictions for each feature column were eliminated by a standardized

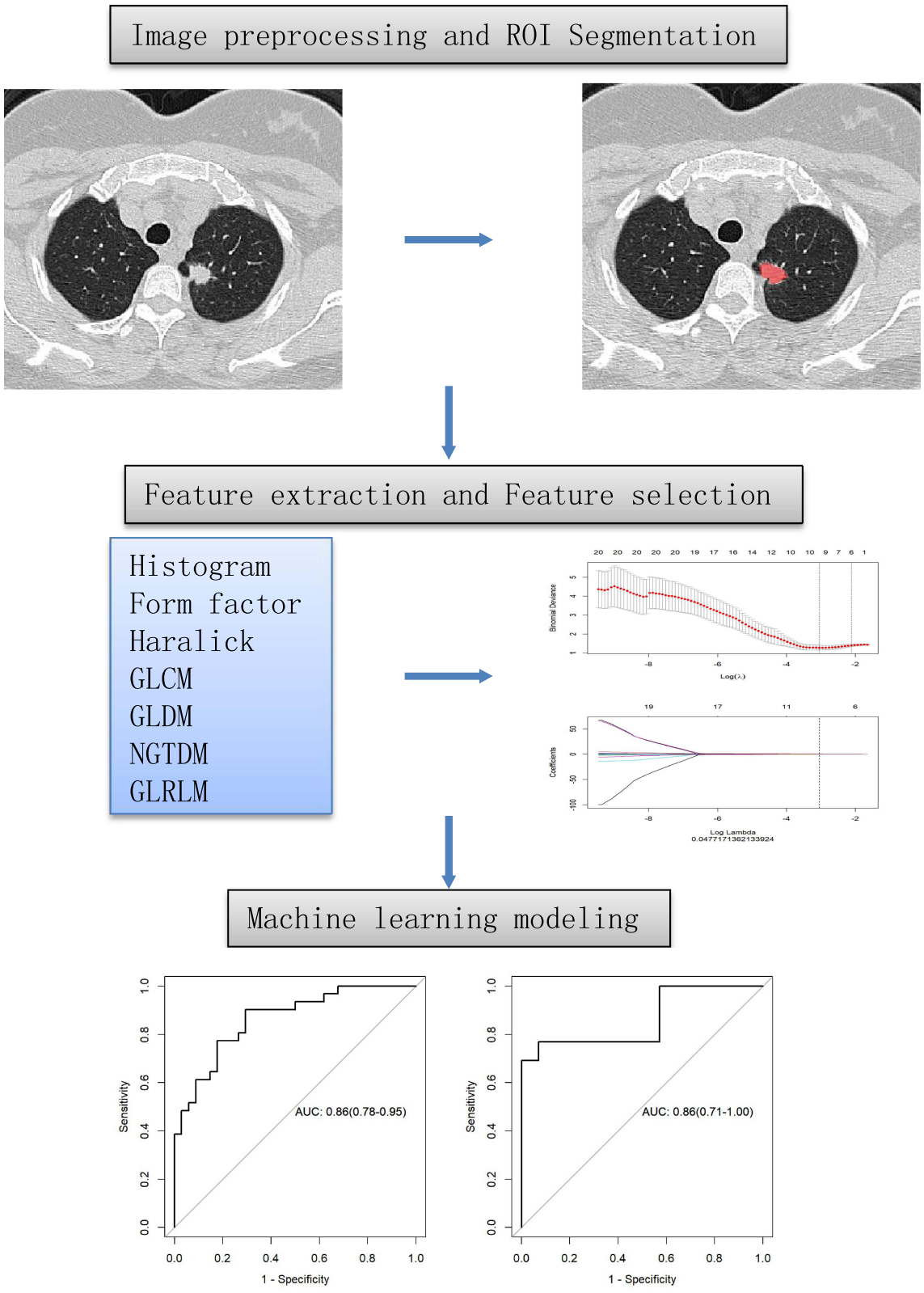

Figure 1 Workflow of radiomics analysis for lung cancer.

approach. Two feature selection algorithms, maximum relevance and minimum redundancy (mRMR) and the least absolute shrinkage and selection operator (LASSO), were used for feature selection. Firstly, mRMR was used to eliminate redundant and irrelevant features. Then, the LASSO regression model was used to reduce the dimension,

and the method of dimensionality reduction was cross-validation. This method is suitable for regression analysis of high-dimensional data, and an L1 penalty term is added to the function of linear regression, so that some regression coefficients become 0 to achieve the effect of dimensionality reduction.

Model Building

First, the training set and test set data are loaded. According to the selected features, the logistic method in machine learning was used to construct the prediction model. Receiver operating characteristic (ROC) curve and accuracy were used to evaluate the performance of the model in the training group and the test group.

Statistical Analysis

The statistical methods used in feature selection and model construction were the statistical methods provided by AK software. SPSS22.0 statistical software was used for data analysis. Measurement data following normal distribution were represented as mean ± SD, and independent sample *t*-test ® was used to evaluate the age of the mutant group and the wild-type group. Count data were analyzed by χ^2 test. $P < 0.05$ was considered statistically significant.

Results

Patient Characteristics

There was no significant difference in gender and age between the mutant group and the wild-type group in clinical baseline data ($P > 0.05$, Table 1). We made a statistical analysis on smoking exposure, comorbidity, vacuole sign and foliation sign of the two groups, and all showed no statistical significance ($P > 0.05$, Table 2). We found that the number of non-smoking exposure in the mutation group was smaller, but still had no statistical significance. In future studies, we will continue to pay attention to the mutation of non-smoking population.

Feature Selection

The original images of 93 cases of pulmonary nodules were imported into AK software together with the segmented images, and a total of 852 radiomic features were obtained. After dimensionality reduction by mRMR, 20 features were screened out, and 10 features were left after further dimensionality reduction by LASSO (Figures 2 and 3).

Table 1 Comparison of General Data Between Mutant Group and Wild-Type Group (x±s)

	Mutant Group	Wild-Type Group	Statistics	*P* value
The number of cases	44	49	NA	NA
Age, mean ± SD, years	70.5±10.1	66.7±13.6	1.54	>0.05
Gender, Male: Female	30:14	26:23	2.21*	>0.05*

Notes: The measurement data were expressed as mean ± standard deviation (SD). Statistical methods: *t*-test and chi-square test (*).

Table 2 Comparison of Smoking Exposure, Comorbidity and Nodule Characteristics Between Mutant Group and Wild-Type Group

Group	Number of Cases	Exposure to Tobacco Smoking Environment	Diabetes	Vocule Sign	Lobulation
Mutant type group	44	21/44 (47.8%)	9/44 (20.5%)	10/44 (22.7%)	36/44 (81.8%)
Wild-type group	49	19/49 (38.8%)	11/49 (22.4%)	15/49 (30.6%)	38/49 (77.6%)
Statistic		χ^2=0.758	χ^2=0.055	χ^2=0.733	χ^2=0.26
P value		>0.05	>0.05	>0.05	>0.05

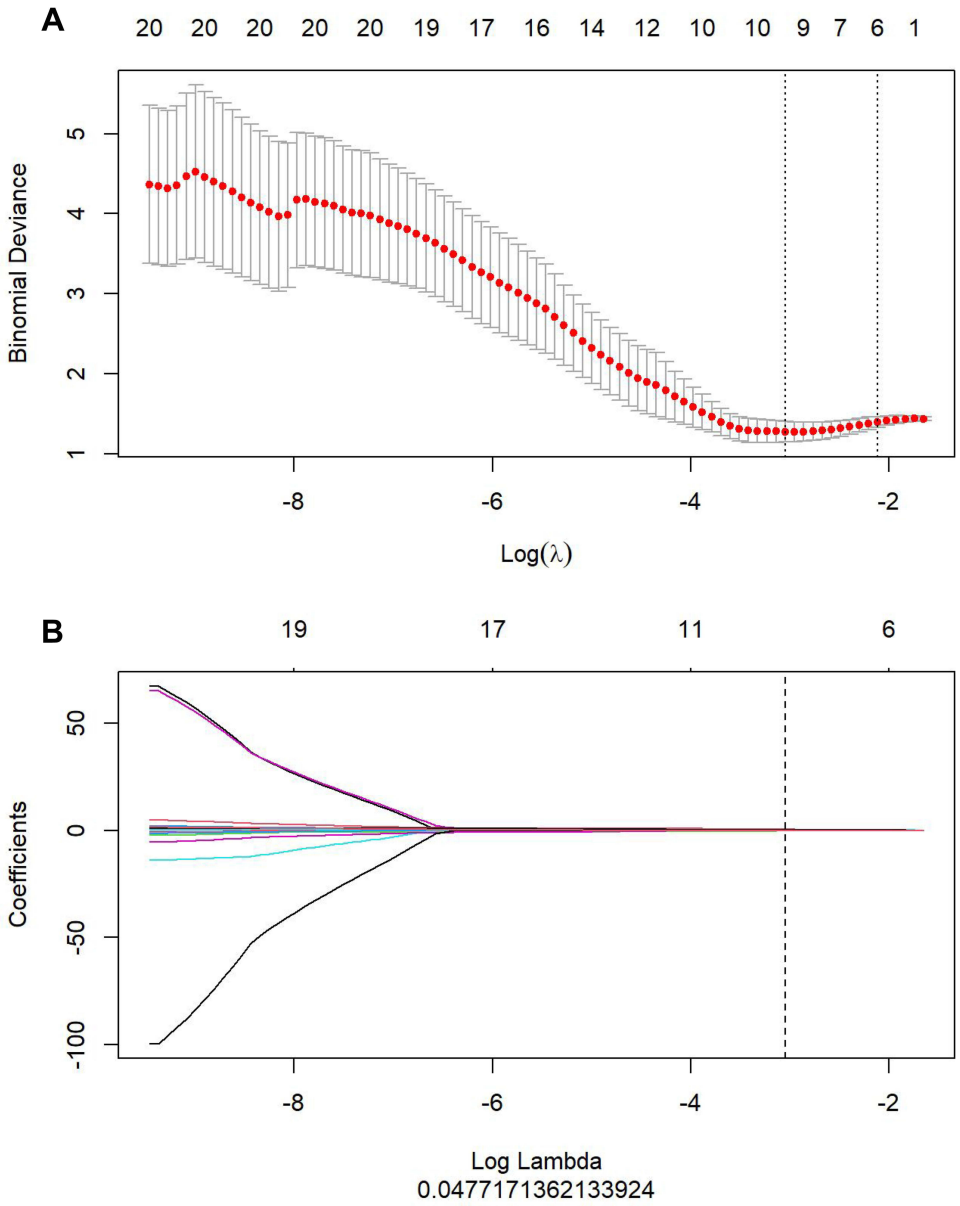

Figure 2 (**A**) Error rate curve. (**B**) LASSO coefficient diagram. We chose the coefficient with the lowest error rate.

Model Building

Based on the 10 selected features, logistic regression was used to construct prediction models in the training and test groups. The accuracy, areas under the curve, specificity, sensitivity, positive predictive value and negative predictive value of the training set were 0.80, 0.86, 0.89, 0.74, 0.90 and 0.71, respectively. The corresponding indicators of the test set were 0.59, 0.86, 0.67, 0.56, 0.77, and 0.43, respectively (Figure 4). The two groups of models fitted well.

Discussion

In this study, AK software was used to extract the features of pulmonary nodules, and it was found that there were characteristic differences in GLCM and RLM between TP53 mutant and wild-type pulmonary nodules, which could be used as an objective indicator for the differentiation of TP53 mutant and wild-type pulmonary nodules.

TP53 gene, a tumor suppressor gene encoding the tumor protein p53, is the most common mutated gene in a variety of human cancers.[14] TP53 is located in the short arm of chromosome 17 (17p13), and exon 5–8 is a hotspot region for mutation

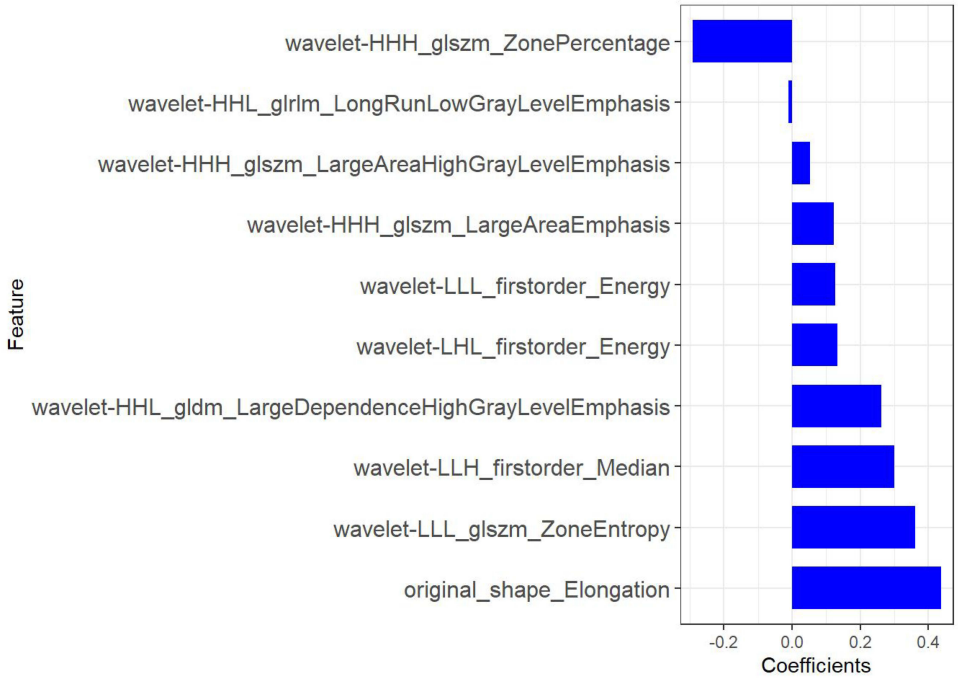

Figure 3 The remaining features after two-step feature dimensionality reduction.

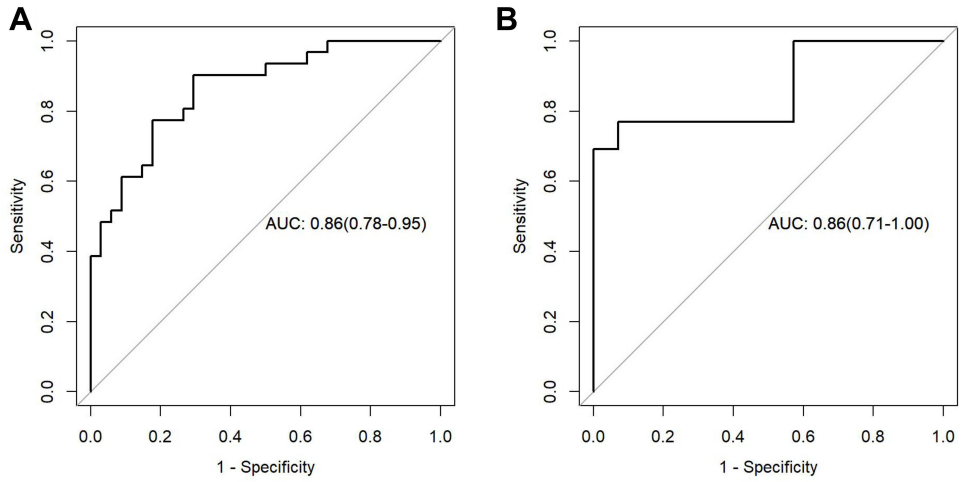

Figure 4 Receiver operating characteristic (ROC) curves of training set (**A**) and test set (**B**).

in human tumors. The DNA-binding domain (DBD) encoding exons 5 through 8 of TP53 consists of 102 to 292 residues. In DBD, the L2 loop (residues 163 to 195) and L3 loop (residues 236 to 251) bind to Zn atoms and play a key role in the interaction with DNA.[15] The p53 protein contains different functional domains: an N-terminal deactivation domain, a sequence-specific DNA-binding region, an oligomerization domain, and a C-terminal negative regulatory domain. Wild-type *p53* gene is stable cell genes and complete "guardian", around normal circumstances, and p53 protein ubiquitin ligase MDM2 combination, and rapid degradation. Therefore, various types of cellular stress cannot be detected in the cells, such as UV-guided DNA damage and oncogene activation blocking MDM2 binding to p53, which then oligomerizes into an active tetramer, allowing the protein to accumulate in the nucleus and transform normal cells into cancer cells.

The occurrence of lung cancer is a complex biological process with multiple factors and multiple stages, which is closely related to oncogene activation and tumor suppressor gene inactivation. With the introduction of next-

generation sequencing technology, it has been found that about 80% of NSCLC progression involves multiple genetic abnormalities, leading to malignant transformation of bronchial epithelial cells, followed by invasion, lymph node metastasis and distant metastasis. Among these genetic abnormalities, *TP53* gene was the most common co-mutation. P53 protein (mutant) is the product of mutations in the *TP53* gene and is a tumor promoter that abolishes normal p53 function. P53 protein encoded by mutated gene in malignant tumors is released from human blood, which can induce the body's autoimmune response, resulting in the production of p53 autoantibodies. Studies have shown that the detection of serum p53 antibody can be used in early diagnosis of lung cancer, even in subclinical stage.[16] Other studies have shown that the positive expression rates of p53 antibody in stage II, IIIa, IIIB and IV lung cancer are very high, respectively, 41.7%, 70.3%, 68% and 73.2%, which has important clinical value for the diagnosis of lung cancer.[17] *P16* as a tumor suppressor gene has been shown to interact with P53 increasing the action of cell cycle regulation. The observation that *p16* gene preferentially undergoes gene silencing through promoter hypermethylation is of crucial importance in the evidence of protein expression in NSCLC.[18,19]

Obtaining accurate molecular phenotypes is the prerequisite for guiding targeted therapy of NSCLC. However, most tumors have high temporal and spatial heterogeneity at the genomic level,[20] and it is sometimes difficult to accurately detect gene mutations in tissue specimens obtained by biopsy. Radiomics is defined as the use of high-throughput to extract features from medical images.[21,22] Associating image features with tumor gene–protein features or tumor phenotypes can develop models for cancer diagnosis, patient prognosis, or relative tumor heterogeneity, especially for patients who are unable to perform biopsy or whose biopsy fails.[23] In the long run, radiomics will not replace the role of radiologists. Instead, it will be a powerful tool for radiologists.[24] In addition, artificial intelligence deep learning on image feature big data can guide clinical decisions[25] and provide valuable information for personalized treatment. Image group learning can in the noninvasive and low-cost way to improve the diagnosis and prognosis of lung cancer and so on, combined with lung form complex, different kinds, strong heterogeneity, applying image omics to lung cancer, more meaningful to clinical decision-making, and provide help for individualized treatment,[26] including lung cancer-related image group study directions in the future more secure, more efficient training pattern of model, Multimodal images are fused and other omics are combined at the same time.[27] The radiomics can well predict the malignancy of solid pulmonary nodules.[28] At home and abroad, there are many radiomics studies on EGFR[29] and ALK[30,31] gene mutations. These studies confirmed the correlation between the expression of these genes and the imaging features, and also laid a theoretical basis for the study in this paper, but there are few radiomics studies on TP53 mutation, we will continue to explore in the following research. In addition, radiomics has also been applied to the study of other diseases, such as Langerhans cell histiocytosis.[32]

However, there are still several limitations in our study. Firstly, the sample size of this study is relatively small. In future studies, we will continue to collect and try to do a large sample study. Secondly, staging and staging methods are not involved in this study. Finally, the lack of consideration of causality is a limitation of this study. The radiomic research on TP53-related lung cancer still needs large-scale multicenter studies.

Conclusion

In conclusion, the radiomic features and classification model of TP53-related lung cancer have certain significance for the evaluation of TP53-related lung cancer. The above radiomics process is convenient and feasible, which is conducive to the diagnosis of patients and the formulation of individualized treatment plans.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81871337), the Zhejiang Provincial Natural Science Foundation of China (No. Y22H185692), the Zhejiang Provincial Medical and Health Technology Project (No. 2020RC092).

Disclosure

Yuguo Wei is affiliated with GE Healthcare Life Sciences. The authors report no other conflicts of interest in this work.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. doi:10.3322/caac.21492
2. Thawani R, McLane M, Beig N, et al. Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer*. 2018;115:34–41. doi:10.1016/j.lungcan.2017.10.015
3. Yu G, Li N, Zhao Y, et al. Salidroside induces apoptosis in human ovarian cancer SKOV3 and A2780 cells through the p53 signaling pathway. *Oncol Lett*. 2018;15(5):6513–6518. doi:10.3892/ol.2018.8090
4. Gyorffy B, Bottai G, Lehmann-Che J, et al. TP53 mutation correlated genes predict the risk of tumor relapse and identify MPS1 as a potential therapeutic kinase in TP53-mutated breast cancers. *Mol Oncol*. 2014;8(3):508–519. doi:10.1016/j.molonc.2013.12.018
5. Gazdar AF, Bunn PA, Minna JD. Small-cell lung cancer: what we know, what we need to know and the path forward. *Nat Rev Cancer*. 2017;17 (12):765. doi:10.1038/nrc.2017.87
6. Yachida S, White CM, Naito Y, et al. Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long term survivors. *Clin Cancer Res*. 2012;18(22):6339–6347. doi:10.1158/1078-0432.CCR-12-1215
7. Schulz-Heddergott R, Ute M. Gain-of-function (Gof) mutant p53 as actionable therapeutic targets. *Cancers*. 2018;10(188):l–16. doi:10.3390/cancers10060188
8. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–446. doi:10.1016/j.ejca.2011.11.036
9. Wu YJ, Liu YC, Liao CY, et al. A comparative study to evaluate CT-based semantic and radiomic features in preoperative diagnosis of invasive pulmonary adenocarcinomas manifesting as subsolid nodules. *Sci Rep*. 2021;11(1):66. doi:10.1038/s41598-020-79690-4
10. Wu YJ, Wu FZ, Yang SC, et al. CH LiangRadiomics in early lung cancer diagnosis: from diagnosis to clinical decision support and education. *Diagnostics*. 2022;12(5):1064. doi:10.3390/diagnostics12051064
11. Park S, Kim B, Lee J, et al. GGO nodule volume-preserving nonrigid lung registration using GLCM texture analysis. *IEEE Trans Biomed Eng*. 2011;58(10):2885–2894. doi:10.1109/TBME.2011.2162330
12. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern*. 1973;SMC-3(6):610–621. doi:10.1109/TSMC.1973.4309314
13. Bocchino C, Carabellese A, Caruso T, et al. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognit Lett*. 1990;11 (6):415–419. doi:10.1016/0167-8655(90)90112-F
14. Wang Y, Wang Z, Piha-Paul S, et al. Outcome analysis of Phase I trial patients with metastatic kras and/ or tp53 mutant non-small cell lung cancers. *Oncotarget*. 2018;9(70):33258–33270. doi:10.18632/oncotarget.25947
15. Molina-Vila MA, Bearan-Alamillo J, Gasco A, et al. Nondisruptive p53 mutations are associated with shorter survival in patients with advanced non-small cell lung cancers. *Clin Cancer Res*. 2014;20(17):4647–4659. doi:10.1158/1078-0432.CCR-13-2391
16. Wan LL, Liang Y, He YT, et al. Value of plasma p16 gene methylation combined with p53 antibody in the diagnosis of non-small cell lung cancer. *J Chin Oncol*. 2017;23(5):372–376.
17. Mattioni M, Soddu S, Prodosmo A, et al. Prognostic role of serum p53 antibodies in lung cancer. *BMC Cancer*. 2015;15:148–159. doi:10.1186/s12885-015-1174-4
18. Pezzuto A, D'Ascanio M, Ricci A, et al. Expression and role of p16 and GLUT1 in malignant diseases and lung cancer: a review. *Thorac Cancer*. 2020;11(11):3060–3070. doi:10.1111/1759-7714.13651
19. Pezzuto A, Cappuzzo F, D'Arcangelo M, et al. Prognostic value of p16 protein in patients with surgically treated non-small cell lung cancer; relationship with Ki-67 and PD-L1. *Anticancer Res*. 2020;40(2):983–990. doi:10.21873/anticanres.14032
20. Schwarz RF, Ng CK, Cooke SL, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS Med*. 2015;12(2):e1001789. doi:10.1371/journal.Pmed.1001789
21. Hatt M, Tixier F, Visvikis D, et al. Radiomics in PET/CT: more than meets the eye? *J Nucl Med*. 2017;58(3):365. doi:10.2967/jnumed.116.184655
22. Kumar V, Gu YH, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imag*. 2012;30(9):1234–1248. doi:10.1016/j.mri.2012.06.010
23. Zhao H, Su Y, Wang M, et al. The machine learning model for distinguishing pathological subtypes of non-small cell lung cancer. *Front Oncol*. 2022;12:875761. doi:10.3389/fonc.2022.875761
24. Lee G, Park H, Bak SH, et al. Radiomics in lung cancer from basic to advanced: current status and future directions. *Korean J Radiol*. 2020;21 (2):159–171. doi:10.3348/kjr.2019.0630
25. Fave X, Zhang L, Yang J, et al. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Scient Rep*. 2017;7 (7):588. doi:10.1038/s41598-017-00665-z
26. Chetan MR, Gleeson FV. Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur Radiol*. 2021;31(2):1049–1058. doi:10.1007/s00330-020-07141-9
27. Wu G, Jochems A, Refaee T, et al. Structural and functional radiomics for lung cancer. *Eur J Nucl Med Mol Imaging*. 2021;48(12):3961–3974. doi:10.1007/s00259-021-05242-1
28. Zhang R, Wei Y, Shi F, et al. The diagnostic and prognostic value of radiomics and deep learning technologies for patients with solid pulmonary nodules in chest CT images. *BMC Cancer*. 2022;22(1):1118. doi:10.1186/s12885-022-10224-z
29. Jia TY, Xiong JF, Li XY, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur Radiol*. 2019;29(9):4742–4750. doi:10.1007/s00330-019-06024-y
30. Jeong CJ, Lee HY, Han J, et al. Role of imaging biomarkers in predicting anaplastic lymphoma kinase-positive lung adenocarcinoma. *Clin Nucl Med*. 2015;40:e34–e39. doi:10.1097/RLU.0000000000000581
31. Le NQK, Kha QH, Nguyen VH, et al. Machine learning-based radiomics signatures for EGFR and KRAS mutations prediction in non-small-cell lung cancer. *Nt J Mol Sci*. 2021;22(17):9254. doi:10.3390/ijms22179254
32. Niu J, Liang J, Feng Q, et al. ^{18}F-FDG PET/MR assessment of pediatric Langerhans cell histiocytosis. *Int J Gen Med*. 2021;14:6251–6259. doi:10.2147/IJGM.S327134