Original Article

Central Asian Journal of Medical Sciences
CAJMS

Cent Asian J Med Sci. 2026 Jun;12(2):58-68 https://doi.org/10.24079/CAJMS.2026.02.006

ALDH18A1 Expression Profiling in Non-small Cell Lung Cancer through Bioinformatics and Experimental Approaches

Wurihan[1,2], Yirigui[3], Jie Cheng[2], Sarantuya Enkhjargal [1], Nyamdorj Dagdanbazar[4], Hanjun Pei[5], Shiirevnyamba Avirmed[6], Uurtuya Shuumarjav[1]

[1]Graduate School of Medical Sciences, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia;
[2]Tumor Center, The First Affiliated Hospital of Baotou Medical College, Inner Mongolia University of Science &Technology, Inner Mongolia, China;
[3]School of Preclinical and Forensic Medicine, Baotou Medical College, Inner Mongolia University of Science & Technology, Inner Mongolia, China;
[4]School of Bio-medicine, Mongolian National University of Medical Sciences, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia;
[5]The Department of Cardiology, The First Affiliated Hospital of Baotou Medical College, Inner Mongolia University of Science & Technology, Inner Mongolia, China;
[6]Department of Surgery, School of Medicine, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia.

Submitted date: Mar 21, 2026
Accepted date: Jun 29, 2026

Corresponding Author:
Uurtuya Shuumarjav (MD., PhD., Assoc Prof.)
Graduate School of Medical Sciences, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia
Email:uurtuya@mnums.edu.mn
ORCID:https://orcid.org/0000-0002-9951-2457
Shiirevnyamba Avirmed (MD., PhD., Prof.)
Department of Surgery, School of Medicine, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia
Email:shiirevnyamba@mnums.edu.mn
ORCID:https://orcid.org/0000-0002-1010-8221

Objective: Non-small cell lung cancer (NSCLC) is a leading cause of cancer-related death worldwide. The molecular mechanisms of NSCLC are still not fully understood. This study used bioinformatics analysis of public databases to identify potential biomarkers and therapeutic targets. This study used integrated bioinformatics analyses to identify candidate oncogenes associated with in NSCLC. Tissue microarray-based immunohistochemistry was then performed to characterize the protein expression pattern of *ALDH18A1* in NSCLC tissues and to preliminarily evaluate its potential as a biomarker. **Methods:** In this study, TCGA transcriptomic data were analyzed by weighted gene co-expression network analysis (WGCNA), differential expression analysis, least absolute shrinkage and selection operator (LASSO) regression, and random forest analysis to identify *ALDH18A1* as a candidate gene in NSCLC. The protein expression of ALDH18A1 was then examined by immunohistochemistry (IHC) using tissue microarrays (TMAs) containing 80 lung squamous cell carcinoma samples, 52 lung adenocarcinoma samples, and their paired adjacent normal tissues. **Results:** In this study, WGCNA, differential expression analysis, LASSO regression, and random forest analysis identified *ALDH18A1* as a key candidate gene in NSCLC. Immunohistochemical analysis showed that the expression level of ALDH18A1 protein was significantly higher in NSCLC tissues than in paired adjacent normal tissues ($p < 0.05$). **Conclusions:** *ALDH18A1* is markedly upregulated in NSCLC and appears to contribute to tumorigenesis and disease progression, highlighting its promise as a potential diagnostic biomarker and a candidate target for therapeutic intervention.

Keywords: Carcinoma, Non-Small-Cell Lung; *ALDH18A1*; Bioinformatics; Gene Expression Profiling; Immunohistochemistry.

Introduction

Malignancies represent a substantial global burden on human health. According to the

Global Cancer Statistics 2020 published by the International Agency for Research on Cancer (IARC), lung cancer accounts for approximately 11.4% of all newly diagnosed cancer cases worldwide, ranking second among all malignancies. It remains the leading cause of cancer-related mortality, contributing to nearly 18% of global cancer deaths.[1] Despite advances in diagnosis and treatment, the overall 5-year survival rate for lung cancer remains below 20%.[2] Lung cancer continues to impose a substantial burden on global public health. The cause of this cancer is still not fully known. Cigarette smoking is recognized as the predominant risk factor, whereas additional contributors, including air pollution, occupational exposures, ionizing radiation, genetic susceptibility, and pre-existing pulmonary diseases also play important roles in lung carcinogenesis.[3] Lung cancer, primarily referring to bronchogenic carcinoma, is a highly heterogeneous malignancy characterized by diverse clinicopathological features. According to histopathological classification, approximately 85% of cases are classified as non-small cell lung cancer (NSCLC), whereas the remaining 15% are categorized as small cell lung cancer (SCLC).[4]

Among the histological subtypes of NSCLC, lung adenocarcinoma (ADC) is the predominant form, followed by lung squamous cell carcinoma (SQCC).[5] Early detection, timely diagnosis, and prompt surgical intervention can markedly improve survival outcomes in patients with lung cancer. However, owing to the insidious onset of NSCLC, only approximately one-third of patients are candidates for surgical resection at the time of diagnosis.[6] Most patients are diagnosed at an advanced or metastatic stage, thereby precluding the opportunity for curative surgical resection.[7] With the rapid progress of precision medicine, the management of NSCLC has expanded beyond conventional surgical resection and chemoradiotherapy. The advent of molecularly targeted therapies and immunotherapeutic strategies has markedly improved patient survival outcomes.[8] However, the absence of actionable therapeutic targets limits the benefit of these strategies in a subset of patients, and the inevitable development of drug resistance further undermines their efficacy. Accordingly, elucidating the molecular mechanisms underlying NSCLC progression and identifying novel oncogenic drivers and therapeutic targets are of critical importance for improving clinical outcomes.

Bioinformatics analysis is an interdisciplinary discipline that applies computational, mathematical, and statistical approaches to the analysis and interpretation of biological data, with a particular focus on molecular-level information.[9] In recent years, the rapid development of high-throughput sequencing technologies, such as RNA sequencing (RNA-seq), chromatin immunoprecipitation sequencing (ChIP-seq), and whole-genome sequencing (WGS), has substantially enhanced the role of bioinformatics in biomedical research, drug discovery, and precision medicine. Bioinformatics data mining generally comprises a series of interconnected steps, including data acquisition, preprocessing, sequence alignment (or mapping), feature extraction, statistical analysis and interpretation, followed by visualization, reporting, and integrative model construction.[10] In recent years, high-throughput gene expression profiling technologies have become increasingly widely applied in biomedical research.[11] Next-generation sequencing (NGS) and microarray-based profiling are widely used as fundamental tools in oncological research. They have broad clinical applications, including molecular tumor classification, prediction of treatment response, prognostic evaluation, molecular diagnosis, identification of novel drug targets, and patient stratification.[12] The integration of gene co-expression network analysis, differential expression analysis, and machine learning-based feature selection has become an effective approach in tumor molecular research, facilitating the systematic identification of cancer-associated genes and potential therapeutic targets.

Integrated bioinformatics analyses have enabled the identification of several key genes involved in tumor initiation and progression. Among them, aldehyde dehydrogenase 18 family member A1 (*ALDH18A1*) is an important proline metabolic enzyme.[13] *ALDH18A1* is an ATP-and-NAD(P)H-dependent mitochondrial enzyme that catalyzes the conversion of glutamate to pyrrolidine-5-carboxylate (*P5C*), which is an intermediate product of proline metabolism and is subsequently converted to proline by *PYCR1*.[14] The human *ALDH* gene family has 19 homologous genes,[15] of which the *ALDH18A1* gene is located on chromosome 10.[16]

In recent years, increasing evidence has indicated that *ALDH18A1* expression is elevated in melanoma tissues and cell lines. Knockdown of *ALDH18A1* has been demonstrated to significantly impair melanoma cell viability and tumor growth.[17] Consistently, *ALDH18A1* silencing markedly suppresses the proliferation of multiple malignancies, including lymphoma, colorectal cancer, breast cancer, and prostate cancer.[18] Moreover, ALDH18A1 is weakly expressed in normal liver tissue but is markedly overexpressed in hepatocellular carcinoma.[19]

ALDH18A1 is upregulated in multiple tumor types, where it contributes to tumor progression and has been proposed as a potential target for anticancer drug development.[20] However, the expression profile and biological significance of *ALDH18A1* in NSCLC have not yet been fully elucidated.

Although considerable efforts have been devoted to elucidating the molecular mechanisms of NSCLC, the molecular basis underlying its development and progression has not been fully understood. Some candidate genes identified through bioinformatics analyses still lack experimental validation, which limits their potential clinical application. Therefore, this study used TCGA database and applied WGCNA and differential expression analysis to identify candidate genes associated with NSCLC. The study further employed tissue microarray-based immunohistochemistry to validate the protein expression pattern of *ALDH18A1* in NSCLC tissues and to preliminarily evaluate its potential as a biomarker. These findings provide a theoretical basis for further investigating the role of *ALDH18A1* in the development and progression of NSCLC and its clinical application value.

Material and Methods

Study Design

This retrospective study integrated bioinformatics analyses and tissue microarray (TMA)-based immunohistochemistry (IHC) to investigate the expression pattern of *ALDH18A1* in NSCLC. The study was conducted in the Department of Pathology, The First Affiliated Hospital of Baotou Medical College, Baotou, Inner Mongolia, China.

Bioinformatics Analysis

TCGA Data Analysis

RNA-seq transcriptomic data of NSCLC were retrieved from the TCGA database (https://portal.gdc.cancer.gov/). All expression profiles were generated using the Illumina HiSeq platform. The dataset included 1,153 samples, consisting of 1,043 tumor and 110 normal lung tissues. Raw count data were obtained in HTSeq-count format. Raw HTSeq-count data were integrated to generate a gene expression matrix and normalized prior to downstream analyses. Differential expression analysis between tumor and normal tissues was conducted using the limma R package. Genes with $p < 0.05$ and $|\log_2$ fold change (logFC)$| > 1$ were identified as DEGs.

Weighted Gene Co-expression Network Analysis (WGCNA)

In this study, WGCNA was performed using the WGCNA R package to identify biologically relevant gene modules associated with NSCLC. The top 5,000 genes ranked by variance were first selected for network construction. The optimal soft-thresholding power (ß) was determined using the "pickSoftThreshold" function in the WGCNA R package by evaluating values ranging from 1 to 20. The scale-free topology fit index (R2) and mean connectivity were calculated for each ß. Following WGCNA criteria, a network is considered to approximate scale-free topology when R2 exceeds 0.85 while maintaining sufficient connectivity. Accordingly, ß = 6 was selected as the optimal soft-thresholding power for subsequent analyses. The adjacency matrix was subsequently converted into a topological overlap matrix (TOM), followed by hierarchical clustering to construct a gene dendrogram. Gene modules were identified based on expression similarity, with distinct branches of the dendrogram representing different modules, each assigned a unique color label. Module-trait correlation analysis was then performed to identify biologically meaningful modules and hub genes associated with disease progression.

Feature Selection Using LASSO Regression and Random Forest

In this study, least absolute shrinkage and selection operator (LASSO) logistic regression and random forest algorithms were employed to identify diagnostic biomarkers for NSCLC. LASSO regression was implemented using the glmnet R package, with feature selection achieved by penalizing regression coefficients. Random forest, an ensemble learning approach based on decision trees, was used to estimate feature importance. Importance was quantified by the percentage increase in mean squared error (% IncMSE), and the top 10 features were retained for further analysis.

Main Reagents and Tissue Specimens

Tissue microarrays (TMAs) used in this study were commercially available products purchased from Shanghai Zhuoli Biotechnology Co., Ltd. The specimens consisted of two TMAs, including 80 cases of lung squamous cell carcinoma (LUSC) and 52 cases of lung adenocarcinoma (LUAD), together with matched adjacent non-tumorous tissues. Reagents included a rabbit anti-

ALDH18A1 polyclonal primary antibody, HRP-conjugated goat anti-rabbit IgG secondary antibody, antibody diluent, 50×EDTA antigen retrieval buffer (pH 8.0), 3% hydrogen peroxide (H_2O_2), and 5-10% normal non-immune serum, all obtained from BBI Life Sciences. The DAB chromogenic detection kit was purchased from Shanghai Sangon Biotech Co., Ltd. Hematoxylin, 1% hydrochloric acid-ethanol differentiation solution, and 1% ammonia bluing solution were obtained from Jiuzhou Berlin Biotechnology Co., Ltd. Xylene and ethanol were purchased from Tianjin Yongda Chemical Reagent Co., Ltd., and neutral resin mounting medium was obtained from Shanghai Biomodel Organism Co., Ltd.

Immunohistochemistry (IHC)

Immunohistochemical staining was performed on 4 μm-thick tissue microarray sections using an indirect immunohistochemistry method. Briefly, sections were baked at 60°C for 20 min, deparaffinized in xylene, and rehydrated through graded ethanol. Antigen retrieval was performed in EDTA buffer under heat-induced conditions. Endogenous peroxidase activity was quenched with 3% hydrogen peroxide, followed by blocking with 5-10% normal non-immune serum. Sections were then incubated with anti-ALDH18A1 primary antibody at 37°C for 40 min, followed by incubation with a biotinylated secondary antibody at 37°C for 30 min. Immunoreactivity was visualized using diaminobenzidine (DAB), with hematoxylin counterstaining of nuclei. Finally, slides were dehydrated, cleared, mounted with neutral resin, and imaged under a light microscope.

Evaluation of Immunohistochemical Staining

ALDH18A1 immunoreactivity was predominantly localized in the cytoplasm and appeared as diffuse brown-yellow granules. Two experienced pathologists independently evaluated the staining results while blinded to the clinicopathological information. Staining intensity was scored on a scale from 0 to 3, and the percentage of positive cells was scored from 0 to 4. The final immunoreactivity score was calculated by multiplying the intensity score by the proportion score. Samples with a score of ≤6 were classified as low expression, whereas samples with a score of >6 were classified as high expression.

Statistical Analysis

Statistical analyses were performed using R software (version 4.2.2) and SPSS (version 21.0). Categorical variables were expressed as counts and percentages. Fisher's exact test or Pearson's chi-square test was used for comparisons between groups when appropriate. Paired comparisons between tumor tissues and matched adjacent tissues were performed using McNemar's test. The exact McNemar test was applied when zero-frequency cells were present. All statistical tests were two-sided, and $p < 0.05$ was considered statistically significant.

Ethical Statement

The Research Ethics Committee of the MNUMS (24/3-06). Commercial tissue microarrays were purchased from Shanghai Zhuoli Biotechnology Co., Ltd. Ethical approval for the use of tissue microarray specimens derived from patients who underwent surgical resection was obtained from the Ethics Committee of Shanghai Zhuoli Biotechnology Co., Ltd., Shanghai, China (Approval No. SHLLS-BA-22101102). The specimens were accompanied by anonymized clinicopathological information. Since the tissue samples were obtained from a commercial source, the investigators could not further verify detailed information regarding the original recruiting hospitals. This study was conducted in accordance with the principles of the Declaration of Helsinki.

Results

WGCNA Network Construction and Key Module Identification

In this study, transcriptome expression data related to NSCLC were retrieved from the TCGA database, comprising a total of 1,153 samples, including 110 normal tissues and 1,043 tumor tissues. Based on these expression profiles, WGCNA was performed to construct gene co-expression networks, aiming to identify regulatory modules associated with the initiation and progression of NSCLC. To construct a scale-free gene co-expression network, the scale-free topology fitting index (R2) and mean connectivity were systematically evaluated across a range of soft-thresholding powers (ß). The results demonstrated that with increasing ß values, R2 gradually increased and plateaued, whereas mean connectivity showed a progressive decline. When ß was set to 6, the R2 value first exceeded 0.9, meeting the criterion for scale-free network topology, while the mean connectivity remained within an acceptable range. Therefore, ß = 6 was selected as the optimal soft-thresholding power for constructing the weighted gene co-expression network (Figure 1A). Based on the TOM, gene modules were identified in NSCLC,

resulting in the detection of five distinct modules (Figure 1B): brown (n = 518), blue (n = 1,155), grey (n = 983), turquoise (n = 1,845), and yellow (n = 499). We further performed module-trait correlation analysis. Based on the absolute value of the correlation coefficient, the MEyellow module showed the strongest association with tumor tissues ($|r|=0.83$, $P=2\times10^{-286}$). Although the correlation was negative, this result still indicates a strong relationship between this module and NSCLC. Therefore, the MEyellow module was selected as a key module for subsequent functional analyses (Figure 1C).

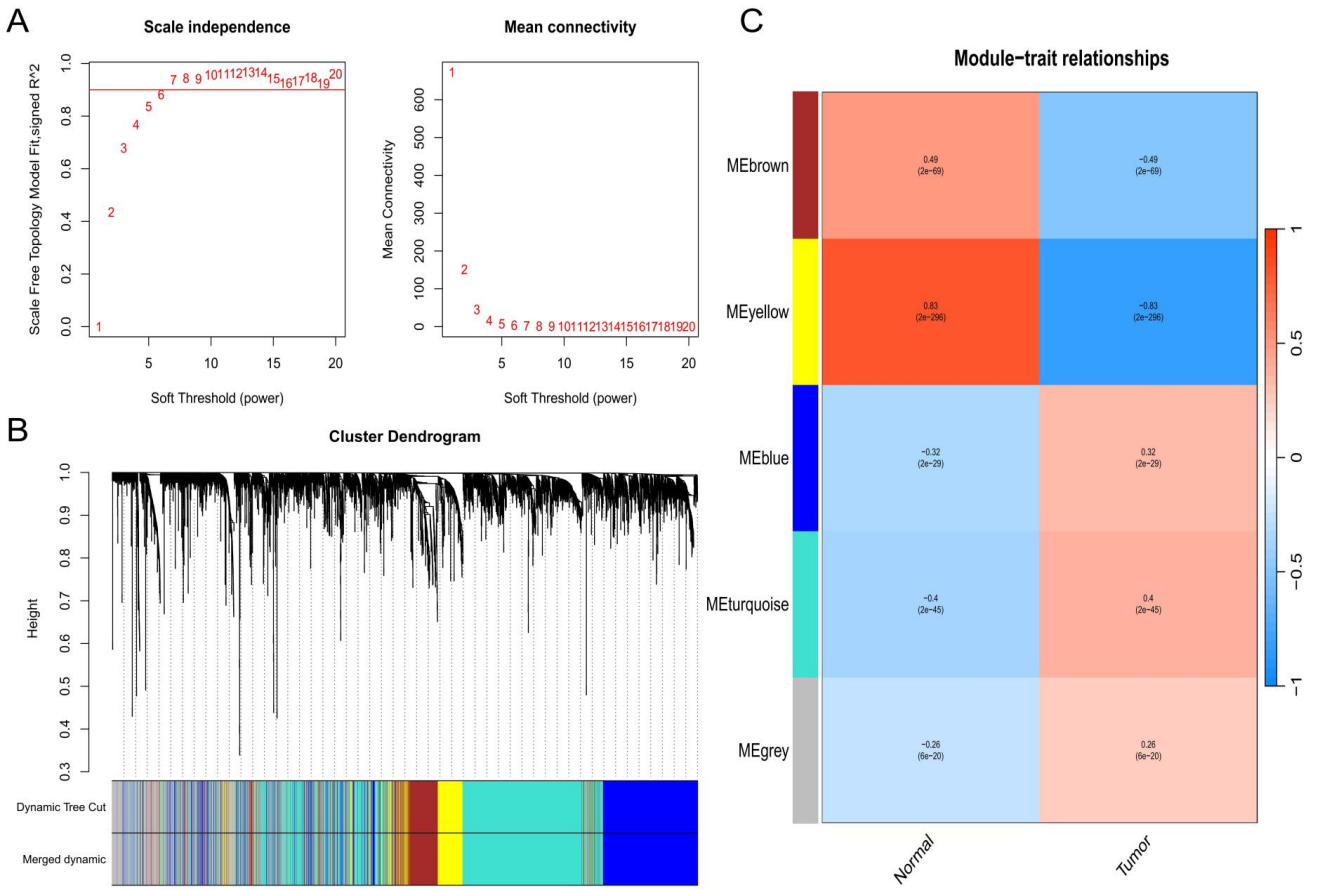

Figure 1. Construction of WGCNA, identification of gene modules, and their association with NSCLC traits.
(A) Determination of the optimal soft-thresholding power for WGCNA. (B) Identification of gene co-expression modules in NSCLC using WGCNA. (C) Module–trait correlation analysis showing significant associations between co-expression modules and clinical phenotypes.

Differentially Expressed Gene Analysis and Intersection with Module Genes

The limma package was used to identify DEGs between the two patient groups. DEGs with $p < 0.05$ and $|\log_2 FC| > 1$ were considered significant. A total of 2402 DEGs were identified, comprising 860 upregulated and 1541 downregulated genes (Figure 2A). The upregulated DEGs were then intersected with the yellow module genes from the previous analysis, yielding 31 overlapping genes (Figure 2B).

DEGs between the two patient groups were identified using the limma package. Genes with $p < 0.05$ and $|\log_2 FC| > 1$ were considered statistically significant. A total of 2,402 DEGs were obtained, including 860 upregulated and 1,541 downregulated genes, with the overall distribution of DEGs shown in Figure 2A. Subsequently, the upregulated DEGs were intersected with genes from the previously identified yellow module, resulting in 31 overlapping genes (Figure 2B).

Feature Gene Selection by LASSO and RF and Identification of Hub Genes

To further identify key genes involved in NSCLC, 31 intersecting genes were subjected to feature selection using LASSO regression and RF analysis. The LASSO model identified 13 genes as potential NSCLC-related feature genes (Figures 3A and 3B). In parallel, RF analysis ranked candidate genes based on importance, and the top 10 genes were selected as NSCLC feature genes (Figure 3C). Subsequently, the genes identified

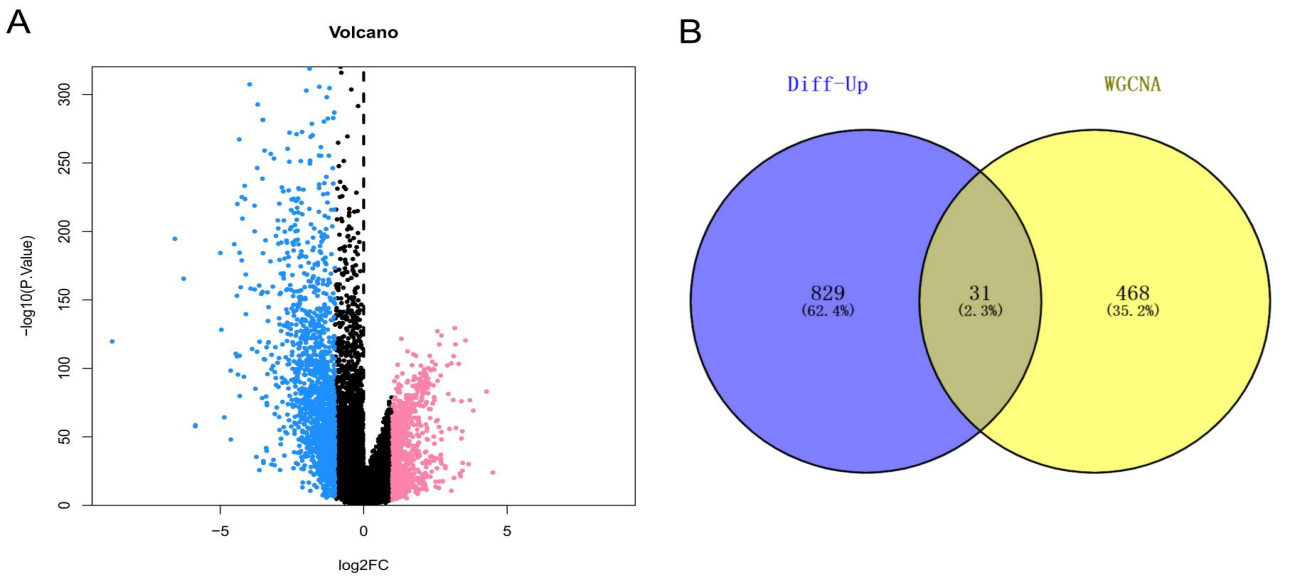

Figure 2. Differential expression analysis and intersection with module genes.
(A) Identification of differentially expressed genes between the two groups, visualized using a volcano plot. (B) Venn diagram illustrating the overlap between significantly upregulated genes and genes within the MEyellow module identified by WGCNA.

by LASSO and RF were intersected, resulting in five overlapping feature genes (Figure 3D). Compared with normal lung tissues, the expression levels of *KAT2A*, *GGCT*, *BZW2*, *ERO1A*, and *ALDH18A1* were significantly upregulated in lung cancer tissues (all $P < 0.001$) (Figure 3E), suggesting that these genes may be involved in the initiation and progression of lung cancer. Among the five candidate genes, *ALDH18A1* showed the most prominent association with tumor-related biological functions based on functional enrichment analysis and published evidence, including regulation of cell cycle progression, maintenance of genomic stability, and modulation of tumor cell invasion and metastasis. Therefore, *ALDH18A1* was preliminarily identified as a core regulatory gene in the development and progression of NSCLC and selected for subsequent validation experiments.

ALDH18A1 Expression in NSCLC and Adjacent Normal Tissues

IHC was performed to evaluate ALDH18A1 protein expression in tumor and adjacent non-tumor tissues from 132 patients with NSCLC. As shown in Table 1, ALDH18A1 expression was detected in 100% of NSCLC tumor tissues, with 99 cases exhibiting high expression and 33 cases showing low expression. Stratified analysis by histological subtype revealed that, among 52 patients with LUAD (LUC1041), 38 cases showed high ALDH18A1 expression (Figure 4A), where ALDH18A1 was predominantly localized in the cytoplasm, while 14 cases exhibited low expression. In 80 patients with LUSC (LUC1603), 61 cases demonstrated high expression (Figure 4C) and 19 cases showed low expression. In contrast, ALDH18A1 expression in adjacent non-tumor tissues (Figures 4B and 4D) was markedly low, with a positivity rate of only 5.3%, including low expression in 3 LUSC-adjacent tissues and 4 LUAD-adjacent tissues. A statistically significant difference in ALDH18A1 protein positivity was observed between tumor tissues and adjacent non-tumor tissues in patients with NSCLC ($p < 0.05$). However, comparison of the high expression rate of ALDH18A1 between LUSC and LUAD revealed no statistically significant difference ($p > 0.05$) (Table 2).

Discussion

Primary lung cancer is the malignant tumor with the highest incidence and mortality worldwide, and it has become one of the leading threats to human health.[21] Lung cancer represents a major focus of research worldwide, with extensive studies conducted both domestically and internationally focusing on its pathogenesis, early diagnosis, therapeutic strategies, and the development of novel antitumor agents. With the improvement of public health awareness and the widespread implementation of lung cancer screenings, therapeutic strategies for lung cancer have continuously evolved, including advances in targeted therapy and immunotherapy.[22] However, the overall 5-year survival rate remains low, and the clinical benefit is still limited.

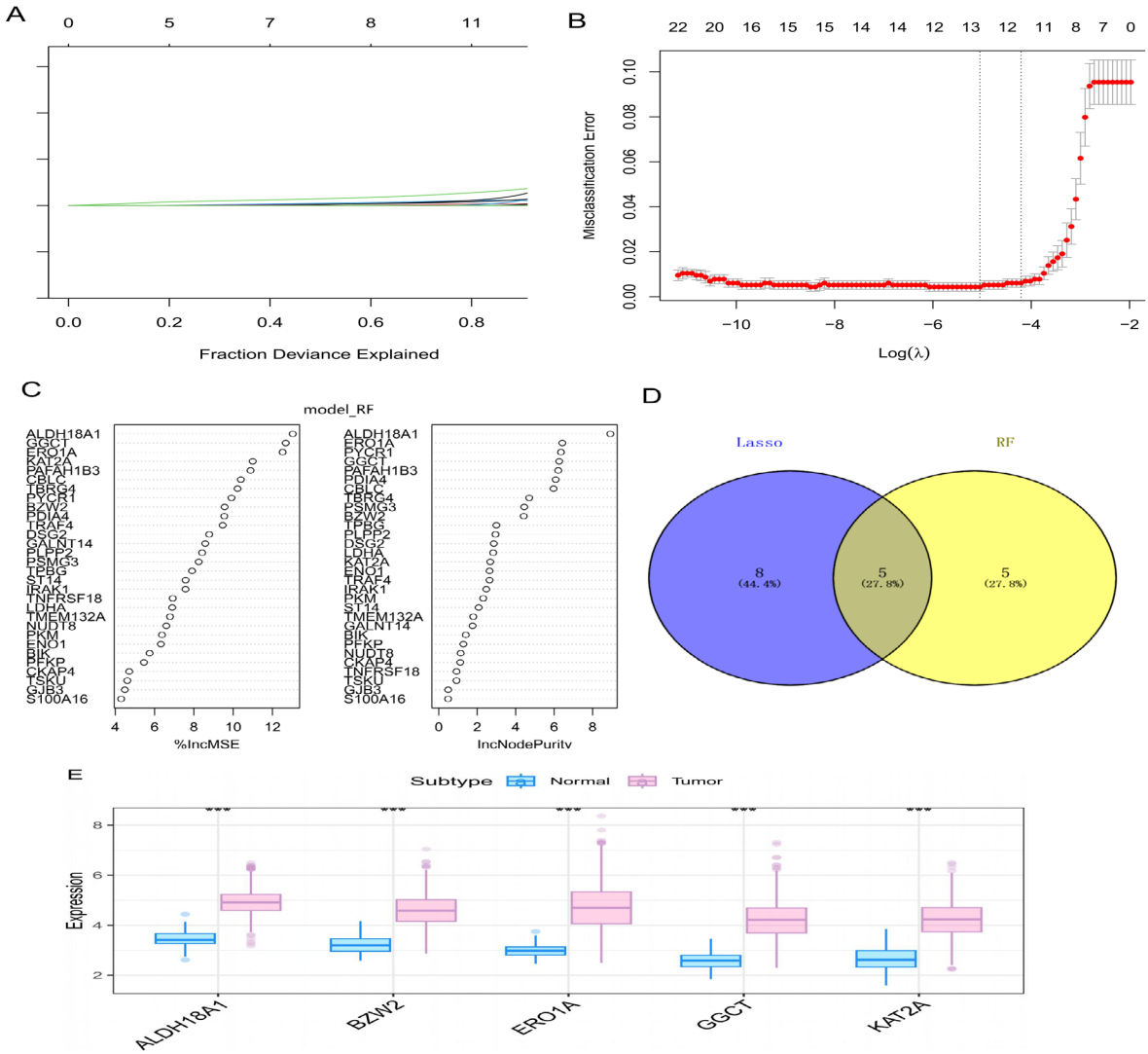

Figure 3.Selection of feature genes using LASSO regression and RF and identification of core genes.
(A) LASSO coefficient profiles of candidate NSCLC genes. (B) Cross-validation curve for selection of the optimal λ in the LASSO model. (C) Ranking of feature importance derived from the random forest model. (D) Venn diagram showing the intersection of feature genes identified by LASSO and RF analyses. (E) Expression levels of core target genes in normal and tumor samples, indicating that *KAT2A, GGCT, BZW2, ERO1A*, and *ALDH18A1* were significantly upregulated in tumor tissues compared with normal tissues. Asterisks (*) denote statistically significant differences between groups ($p < 0.001$).

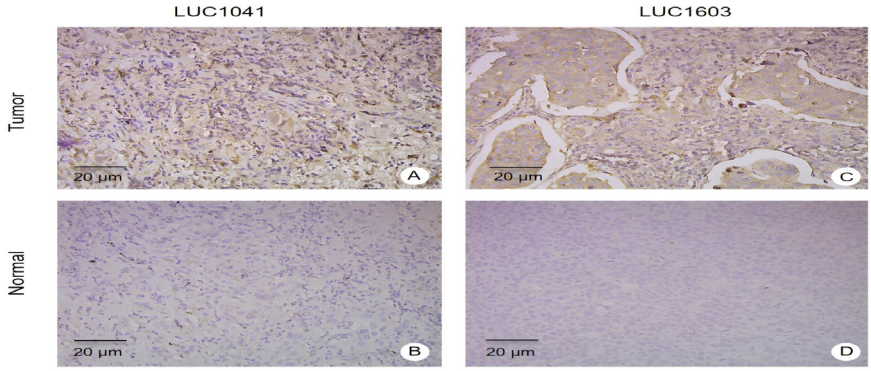

Figure 4. Immunohistochemical staining showing ALDH18A1 positivity in NSCLC tissues and the corresponding adjacent non-tumor tissues.
(A) Lung adenocarcinoma (LUAD); (B) adjacent non-tumor tissue from LUAD; (C) lung squamous cell carcinoma (LUSC); (D) adjacent non-tumor tissue from LUSC. ×200.

Table1. Expression of ALDH18A1 in NSCLC and adjacent tissues

Group	ALDH18A1			Total	Rate (%)	*p*
	High	Low	Negative			
NSCLC Tissues	99	33	0	132	100.0	<0.001
LUSC	61	19	0	80		
LUAD	38	14	0	52		
NSCLC Adjacent	0	7	125	132	5.3	
LUSC	0	3	77	80		
LUAD	0	4	48	52		

Data are presented as number of cases. p values were calculated using McNemar's test.

Table 2. High expression rate of ALDH18A1 in LUSC and LUAD

Group	ALDH18A1		Total	High rate (%)	X^2	*p*
	High	Low				
LUSC	61	19	80	76.25	0.17	0.68
LUAD	38	14	52	73.08		
Total	99	33	132	75.00		

Postoperative recurrence after surgery or active treatment, as well as the emergence of various forms of treatment resistance, continue to pose major challenges for clinicians.[23]

Using bioinformatics analyses based on GEO and TCGA datasets, hub genes associated with tumor initiation, progression, and prognosis have been identified, and regulatory networks have been constructed. These studies highlight the value of large-scale databases in uncovering key tumor-related genes, which contributes to a better understanding of cancer mechanisms and improves prognostic assessment.[24] In recent years, gene microarray and sequencing data have provided a convenient and powerful resource for the identification of potential gene targets in NSCLC.[25] Although multiple candidate genes were identified through bioinformatics screening in this study, *ALDH18A1* was prioritized for further investigation due to its pivotal role in metabolic reprogramming and its well-established involvement in tumorigenesis.

Metabolic reprogramming is a common phenomenon in various diseases, that involves alterations in glucose, lipid, and amino acid metabolism, and is closely associated with disease initiation and progression.[26] Among these, amino acid metabolism has long been a major focus in cancer research and therapeutic development. Recently, it has been demonstrated that proline metabolism plays a critical role in cellular metabolic reprogramming and contributes to the development of multiple cancers, including lung cancer.[27] The proline metabolic pathway encompasses both proline biosynthesis and catabolism. The key enzymes involved in proline metabolism and synthesis are *ALDH18A1* and *PYCR*.[28] *ALDH18A1* is also known as *GSAS, P5CS, PYCS*, and *SPG9*. The cytoplasmic expression of the *ALDH18A1* gene is found in several tissues. Subcellular localization analysis indicates that it is primarily associated with mitochondria and is generally distributed within the intracellular compartment. This gene is a member of the acetaldehyde dehydrogenase family and encodes a bifunctional ATP and NADPH-dependent mitochondrial enzyme with γ-glutamyl kinase and γ-glutamyl phosphate reductase activities.

Aberrant overexpression of *ALDH18A1* has been reported in multiple solid tumors, including esophageal cancer,[29] hepatocellular carcinoma, colorectal cancer, and breast cancer.[30] It is frequently associated with enhanced proliferation, increased invasion, genomic instability, and poor prognosis. We employed TMA technology to evaluate ALDH18A1 protein expression in NSCLC tissues. The results demonstrated that ALDH18A1 expression was significantly elevated in NSCLC tissues, with a higher rate of aberrant expression compared with adjacent normal lung tissues. These findings suggest that *ALDH18A1* may play a role in lung tumorigenesis and that the extent of

its aberrant expression may serve as a potential diagnostic biomarker for cancer.

The innovation of this study lies in the integration of bioinformatics analyses with histopathological validation, establishing a comprehensive workflow from large-scale data mining to tissue-level confirmation, thereby improving the robustness and reliability of the findings. However, several limitations should be acknowledged. First, the tissue microarray samples lacked complete clinical follow-up information, preventing further evaluation of the association between *ALDH18A1* expression and patient prognosis. Second, the data were primarily derived from public databases and require validation in larger, multicenter cohorts. In addition, in vitro and in vivo experiments were not performed to further investigate the biological functions of *ALDH18A1* and its underlying molecular mechanisms.

Future studies should further explore multiple aspects. First, it is necessary to establish prospective or retrospective clinical cohorts with complete follow-up information and to systematically evaluate the association between ALDH18A1 expression and overall survival (OS), progression-free survival (PFS), and clinicopathological characteristics using Kaplan–Meier survival analysis and Cox proportional hazards regression models, thereby clarifying its prognostic value in clinical assessment. Second, at the mechanistic level, in vitro cellular experiments and in vivo animal models (e.g., gene knockdown or overexpression assays, as well as cell proliferation, migration, and invasion assays) are warranted to further validate the biological functions of *ALDH18A1*. Integrative analyses using transcriptomics or proteomics should also be performed to elucidate the key signaling pathways potentially involved in the initiation and progression of NSCLC. In addition, from a translational perspective, future studies may further evaluate the sensitivity and specificity of *ALDH18A*1 as a potential biomarker for early diagnosis and explore its feasibility as a therapeutic target, for instance, by integrating it with existing targeted treatment strategies or developing novel inhibitors for functional validation.

Conclusion

This study showed that ALDH18A1 was significantly overexpressed in non-small cell lung cancer tissues, suggesting that it may serve as a potential molecular biomarker for NSCLC. These findings provide a basis for further studies investigating the role of *ALDH18A1* in NSCLC development and progression and its potential clinical value.

Conflict of Interest

The authors state no conflict of interest.

Authors Contribution

Wurihan: ORCID https://orcid.org/0000-0002-9951-2457, Conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing-original draft, writing-review and editing.

Yirigui: Data curation, investigation, validation.

Jie Cheng : ORCID https://orcid.org/0000-0002-2977-8671, Formal analyses, software, visualization.

Sarantuya Enkhjargal: ORCID https://orcid.org/0000-0003-4730-1110, Investiogation

Nyamdorj Dagdanbazar: ORCID https://orcid.org/0000-0002-7646-6865, Formal analyses, investigation.

Hanjun Pei: ORCID https://orcid.org/0009-0009-4712-0150, Conceptualization, supervision, writing-review and editing.

Shiirevnyamba Avirmed: ORCID https://orcid.org/0000-0002-1010-8221, Conceptualization, supervision, writing-review and editing.

Uurtuya Shuumarjav: ORCID https://orcid.org/0000-0002-9951-2457, Conceptualization, methodology, project administration, supervision, writing-review and editing.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. https://doi.org/10.3322/caac.21660
2. Bray F, Laversanne M, Weiderpass E, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-263. https://doi.org/10.3322/caac.21834
3. Bade BC, Dela Cruz CS. Lung cancer 2020: epidemiology, etiology, and prevention. *Clin Chest Med*. 2020;41(1):1-24.

https://doi.org/10.1016/j.ccm.2019.10.001
4. Siegel RL, Kratzer TB, Giaquinto AN, et al. Cancer statistics, 2025. *CA Cancer J Clin*. 2025;75(1):10-45. https://doi.org/10.3322/caac.21871
5. Thai AA, Solomon BJ, Sequist LV, et al. Lung cancer. *Lancet*. 2021;398(10299):535-554. https://doi.org/10.1016/S0140-6736(21)00317-0
6. Chan LW, Ding T, Shao H, et al. Augmented features synergize radiomics in postoperative survival prediction and adjuvant therapy recommendation for non-small cell lung cancer. *Front Oncol*. 2022;12:659096. https://doi.org/10.3389/fonc.2022.659096
7. Xiong Q, Qin B, Xin L, et al. Real-world efficacy and safety of anlotinib with and without immunotherapy in advanced non-small cell lung cancer. *Front Oncol*. 2021;11:659380. https://doi.org/10.3389/fonc.2021.659380
8. Jeon H, Wang S, Song J, et al. Update 2025: Management of non-small-cell lung cancer. *Lung*. 2025;203(1):53. https://doi.org/10.1007/s00408-024-00700-9
9. Zhou Y, Li H, Chen Y, et al. Comprehensive analysis of differential gene expression to identify common gene signatures in multiple cancers. *Biomed Res Int*. 2021;2021:6616700. https://doi.org/10.1155/2021/6616700
10. Perkel JM. Data visualization tools drive interactivity and reproducibility in online publishing. *Nature.* 2020;587(7832):147-149. https://doi.org/10.1038/d41586-020-02831-4
11. Tzec-Interián JA, González-Padilla D, Góngora-Castillo EB. Bioinformatics perspectives on transcriptomics: a comprehensive review of bulk and single-cell RNA sequencing analyses. *Quant Biol*. 2025;13(2):e78. https://doi.org/10.1007/s40484-025-00378-1
12. Isaic A, Motofelea N, Hoinoiu T, et al. Next-generation sequencing: a review of its transformative impact on cancer diagnosis, treatment, and resistance management. *Diagnostics (Basel*). 2025;15(19):2425. https://doi.org/10.3390/diagnostics15192425
13. Geng P, Qin W, Xu G. Proline metabolism in cancer. *Amino Acids*. 2021;53(12):1769-1777. https://doi.org/10.1007/s00726-021-03046-5
14. Xu X, Zhang G, Chen Y, et al. Can proline dehydrogenase a key enzyme involved in proline metabolism be a novel target for cancer therapy? *Front Oncol*. 2023;13:1254439. https://doi.org/10.3389/fonc.2023.1254439
15. Marchitti SA, Orlicky DJ, Vasiliou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin Drug Metab Toxicol*. 2008;4(6):697-720. https://doi.org/10.1517/17425255.4.6.697
16. Liu G, Snetsinger S, De Jong PJ, et al. Assignment of the human gene encoding the delta1-pyrroline-5-carboxylate synthetase (P5CS) to 10q24.3 by in situ hybridization. *Genomics*. 1996;37(1):145-146. https://doi.org/10.1006/geno.1996.0396
17. Kardos GR, Wastyk HC, Robertson GP. Disruption of Proline Synthesis in Melanoma Inhibits Protein Production Mediated by the GCN2 Pathway. *Mol Cancer Res*. 2017;15(3):281-293. https://doi.org/10.1158/1541-7786.MCR-16-0118
18. Phang JM. Proline metabolism in cell regulation and cancer biology: recent advances and hypotheses. *Antioxid Redox Signal*. 2019;30(4):635-649. https://doi.org/10.1089/ars.2017.7342
19. Boroughs LK, DeBerardinis RJ. Metabolic pathways promoting cancer cell survival and growth. *Nat Cell Biol.* 2015;17(4):351-359. https://doi.org/10.1038/ncb3124
20. Xia J, Chen J, Wang H, et al. Aldehyde dehydrogenase in solid tumors and other diseases. *MedComm (2020)*. 2023;4(1):e195. https://doi.org/10.1002/mco2.195
21. Bray F, Laversanne M, Weiderpass E. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*. 2021;127(16):3029-3030. https://doi.org/10.1002/cncr.33687
22. Phang JM, Liu W, Hancock C, Christian KJ. The proline regulatory axis and cancer. *Curr Opin Clin Nutr Metab Care.* 2015;18(1):71–77. https://doi.org/10.1097/MCO.0000000000000122
23. Barr T, Ma SB, Li ZX, et al. Recent advances and remaining challenges in lung cancer therapy. *Chin Med J (Engl)*. 2024;137(5):533-546. https://doi.org/10.1097/CM9.0000000000002991
24. Wei G, Dong Y, He Z, et al. Identification of hub genes and construction of an mRNA-miRNA-lncRNA regulatory network in gastric carcinoma using integrated bioinformatics analysis. *PLoS One*. 2021;16(12):e0261728. https://doi.org/10.1371/journal.pone.0261728
25. Zhou X, Hao M, Guo X, et al. Identification and validation of differentially expressed genes for non-small cell lung cancer using multiple GEO microarray datasets. *Front*

Oncol. 2023;13:1206768. https://doi.org/10.3389/fonc.2023.1206768

26. Zhu S, Zhao J, Li X, et al. Regulation of glucose, fatty acid and amino acid metabolism by ubiquitination and SUMOylation in disease. *Front Cell Dev Biol*. 2022;10:849625. https://doi.org/10.3389/fcell.2022.849625
27. Daniello C, Patriarca E, Phang J, et al. Proline metabolism in tumor growth and metastatic progression. *Front Oncol.* 2020;10:776. https://doi.org/10.3389/fonc. 2020.00776
28. Linder SJ, Bernasocchi T, Martínez-Pastor B, et al. Inhibition of the proline metabolism rate-limiting enzyme P5CS allows proliferation of glutamine-restricted cancer cells. *Nat Metab*. 2023;5(12):2131-2147. https://doi.org/10.1038/s42255-023-00819-3
29. Zhang Y, Li X, Wang Y, et al. ALDH18A1 has carcinogenic functions and regulates alternative splicing events of DNA repair-related genes in esophageal carcinoma cells. *Sci Rep*. 2022;12(1):10203. https://doi.org/10.1038/s41598-022-14337-1
30. Duan JJ, Wang J, Wang D, et al. Glutamine-arginine-proline axis: a key node in cancer metabolism. *Front Oncol.* 2022;12:963444. https://doi.org/10.3389/fonc.2022.963444