Analysis of HSPD1 Gene Expression and Its Clinical Significance in Lung Adenocarcinoma Through Bioinformatics

Chunling Wang*
Beijing Anzhen Nanchong Hospital, Capital Medical University & Nanchong Central Hospital, Nanchong, China

Abstract: Objective: To clarify the expression, clinical significance, and molecular mechanism of HSPD1 in lung adenocarcinoma (LUAD), and to establish a foundation for predictive evaluation and precision treatment of LUAD. Methods: Based on the TCGA and GEPIA databases, the expression differences of HSPD1 in LUAD were analyzed; combined with clinical information, the correlation between HSPD1 expression and clinical features of LUAD patients such as tumor stage and race was explored; using the Kaplan-Meier Plotter database, the impact of HSPD1 on the overall survival and first progression-free survival of patients was evaluated; through the TIMER2.0 database, the association between HSPD1 expression, copy number variation, and tumor immune cell infiltration was analyzed; with the help of the GeneMANIA database, core interacting genes of HSPD1 were screened, and GO, KEGG enrichment analysis was performed using the R language clusterProfiler package to analyze its molecular regulation pathways. Results: HSPD1 was significantly highly expressed in LUAD and various tumors, and the consistency was verified at both transcriptomic and proteomic levels (P<0.05); its expression increased with the progression of LUAD stage (P<0.05), and was significantly associated with patient race, gender, and age (P<0.05). Patients with high HSPD1 expression had shorter overall survival and increased risk of disease progression (P<0.05), and its expression showed a significant inverse relationship with B cell and CD4+ T cell infiltration (P<0.05), and copy number amplification could further reduce immune cell infiltration. 20 core HSPD1 interacting genes were screened, which are mainly involved in amino acid metabolism, protein folding, and other processes and related pathways. Conclusion: HSPD1 is highly expressed in LUAD, closely related to tumor stage and patient prognosis, and may participate in the progression of LUAD by regulating immune cell infiltration and amino acid metabolism pathways. This molecule may serve as a diagnostic indicator and treatment focus for lung adenocarcinoma.

Keywords: HSPD1; Lung Adenocarcinoma; Bioinformatics; Prognosis; Diagnosis; Biomarkers

1. Introduction

Lung adenocarcinoma (LUAD), as the predominant subtype of non-small cell lung cancer, accounts for approximately 40%-50% of the total incidence of lung cancer. Its onset is insidious and progression is rapid, with the majority of patients diagnosed at advanced stages, leading to a poor prognosis [1-2]. While groundbreaking treatments like targeted therapy and immunotherapy have given certain LUAD patients a new lease on life in recent years, boosting their chances of survival considerably [3], issues like tumor heterogeneity and drug resistance continue to severely limit therapeutic efficacy. Therefore, there is an urgent need to identify prognostic biomarkers and potential therapeutic targets with clinical value to guide improved diagnostic and therapeutic approaches for LUAD.

Heat shock proteins (HSPs) are a class of conserved proteins that are highly expressed under cellular stress conditions. By participating in processes such as protein folding, stabilization, and degradation, they regulate key biological functions including cell proliferation, apoptosis, and metabolism. Altered expression correlates with tumor formation, progression, and outcomes [4]. Heat shock protein family D member 1 (HSPD1), also known as HSP60, serves as a core molecular chaperone in the mitochondrial matrix. This compound not only serves as a linchpin in preserving mitochondrial

equilibrium but also actively fuels cancer advancement by controlling cellular division and preventing programmed cell death [5]. Previous studies have confirmed that HSPD1 is highly expressed in various malignant tumors, including colorectal cancer and liver cancer, and is associated with poor patient prognosis [6]. However, its expression pattern, clinical significance, and molecular regulatory mechanisms in LUAD have not yet been systematically elucidated.

Tumorigenesis emerges from the combined effects of numerous genetic elements and signaling pathways, and the interactions between genes and the pathway regulatory network are the core of analyzing tumor molecular mechanisms [7]. At the same time, the remodeling of the tumor immune microenvironment has become a hotspot in tumor research, and the immune cell infiltration status is closely related to tumor progression, treatment response, and patient prognosis [8]. In addition, as a common type of variation in the tumor genome, gene copy number variations can affect gene expression, thereby regulating the tumor immune microenvironment and participating in the malignant progression of tumors [9]. Therefore, systematically investigating the expression characteristics of HSPD1 in LUAD, its association with clinicopathological parameters and prognosis, its interaction with tumor immune infiltration and copy number variations, and clarifying its molecular interaction network and functional pathways are of great significance for revealing the biological function and clinical value of HSPD1 in LUAD.

This study integrates multi-omics data and clinical information from public databases such as TCGA, GEPIA, and Kaplan-Meier Plotter, employing bioinformatics analysis methods to systematically analyze the expression differences of HSPD1 in pan-cancer and LUAD, to explore its association with the clinical features and survival prognosis of LUAD patients, to elucidate its relationship with immune cell infiltration and copy number variations, and to reveal its potential regulatory mechanisms through molecular interaction network and functional enrichment analysis, aiming to provide new theoretical basis and experimental support for the prognostic assessment and targeted therapy of LUAD.

2. Materials and Methods

2.1 Data Source

Clinical and gene expression data associated with LUAD were retrieved from The Cancer Genome Atlas (TCGA) database [10]. This dataset included a total of 59 normal lung tissue samples and 515 lung adenocarcinoma tissue samples. Subsequently, quality control was performed on the raw data, excluding samples with incomplete case information or missing gene expression data. Finally, qualified samples were selected to statistically analyze the gene expression level of HSPD1 in LUAD.

2.2 Expression Analysis

Utilizing the Tumor Immune Evaluation Resource 2.0 (TIMER 2.0) database [11], transcriptome data of cancerous tissues and paired para-cancerous normal tissues from 22 types of malignant tumors were extracted for statistical analysis of HSPD1 expression levels. Concurrently, the Gene Expression Profiling Interactive Analysis (GEPIA) database [12] was employed, incorporating 483 cases of lung adenocarcinoma tissue samples and 59 cases of normal lung tissue samples to investigate the expression differences of HSPD1 between lung adenocarcinoma and normal lung tissues. Furthermore, the UALCAN database [13] was utilized to mine lung adenocarcinoma-related transcriptome data and clinical information, comparing the expression characteristics of HSPD1 between normal lung tissues and lung adenocarcinoma tissues. Additionally, based on stratification factors such as cancer clinical stage, race, gender, and age, the expression distribution patterns of HSPD1 in different subgroups were further analyzed.

2.3 Survival and Prognosis Analysis

Using the Kaplan-Meier Plotter database [14], obtain the survival follow-up data related to HSPD1 expression. Categorize the samples into two groups-low and high-by comparing HSPD1 expression levels to the median threshold. Draw Kaplan-Meier survival curves for Overall Survival (OS) and First Progression Survival (FPS), respectively. Use the log-rank test to compare the survival differences between the two groups, while simultaneously calculate the Hazard Ratio (HR) and 95% Confidence Interval (CI) to evaluate the strength of the association between HSPD1 expression levels and patient

survival prognosis.

2.4 Immune Infiltration and Copy Number Variation Analysis

Using the TIMER2.0 database [11], we obtained HSPD1 expression data, immune cell (B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, dendritic cells) infiltration level data, and HSPD1 copy number variation information (arm-level deletion, diploid/normal, arm-level amplification, high amplification) in lung adenocarcinoma (LUAD) samples. We calculated the partial correlation coefficients (partial.cor) between HSPD1 expression levels and the infiltration levels of each immune cell type using Spearman partial correlation analysis and performed significance tests. Concurrently, scatter plots and fitting curves were generated to visualize their correlation. Furthermore, we analyzed the distribution differences of immune cell infiltration levels across different HSPD1 copy number variation subtypes, assessed intergroup differences through statistical tests, and presented the results using box plots.

2.5 Construction of the Molecular Network for HSPD1 Interacting Genes

Use the GeneMANIA database [15] to retrieve potential interacting genes of HSPD1, filter out the top 20 interacting genes based on relevance, construct a molecular interaction network of HSPD1 with these 20 genes, and visualize the interaction relationships between the genes.

2.6 GO and KEGG Enrichment Analysis of HSPD1 Interacting Genes

Using the top 20 HSPD1 interacting genes screened from the GeneMANIA database as the research subject, the clusterProfiler package was called in the R language environment to perform GO functional annotation and KEGG pathway enrichment analysis respectively. The GO analysis covers the classifications of Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), while the KEGG analysis targets metabolic and signaling pathways. Finally, the enrichment results are presented in a bubble plot, where bubble size corresponds to gene count and color corresponds to the P value.

3. Result

3.1 Expression Characteristics of HSPD1 in Pan-cancer and LUAD

By analyzing pan-cancer samples from the TCGA database, we found that the expression pattern of HSPD1 is consistent across most tumors: as shown in Figure 1A, the expression level of HSPD1 is significantly higher in multiple tumor tissues, including lung adenocarcinoma (LUAD), breast cancer (BRCA), etc., compared to the corresponding adjacent normal tissues (all $P<0.05$), suggesting that HSPD1 generally exhibits a high-expression trend in tumors.

The specific analysis for lung adenocarcinoma (LUAD) shows (Figure 1B): in 483 tumor tissues and 59 normal lung tissues from the TCGA-LUAD cohort, the expression level of HSPD1 in tumor tissues is significantly higher than in normal tissues ($P<0.05$).

To further verify the reliability of this difference, we conducted validation at both the transcriptome and proteome levels: TCGA transcriptome data show that the transcript level of HSPD1 demonstrates elevated expression in LUAD tumor tissues relative to healthy counterparts (Figure 1C, $P<0.05$); meanwhile, CPTAC proteome data further confirm that the protein expression level of HSPD1 is also significantly higher in LUAD tumor tissues than in normal lung tissues (Figure 1D, $P<0.05$).

3.2 Association of HSPD1 Expression with Clinical Characteristics in LUAD

We also examined HSPD1 expression correlations with LUAD clinical characteristics: In terms of tumor stage (Figure 2A), the transcript levels of HSPD1 progressively increased with the advancement of LUAD stage. The expression levels in patients from Stage 1 to Stage 4 were all significantly higher than in normal tissues (all $P<0.05$), and the expression level in Stage 4 patients was the highest among all stages, suggesting that its expression may be related to the degree of tumor progression; In the patient racial stratification (Figure 2B), the HSPD1 expression levels in Caucasian, African American, and Asian LUAD patients were all significantly higher than in normal tissues (all $P<0.05$), however, expression showed no significant variation across racial groups, indicating that this high expression characteristic is common across different races. From the perspective of gender (Figure 2C), the HSPD1 expression in both male and female LUAD patients was significantly higher than in normal

tissues (all P<0.05), and there was no significant difference between male and female patients, suggesting that its high expression is not related to gender; In the age stratification (Figure 2D), the HSPD1 expression in LUAD patients in all age groups from 21-40 years to 81-100 years was significantly higher than in normal tissues (all P<0.05), and no significant difference was observed among different age groups, indicating that this expression characteristic is not affected by the patient's age.

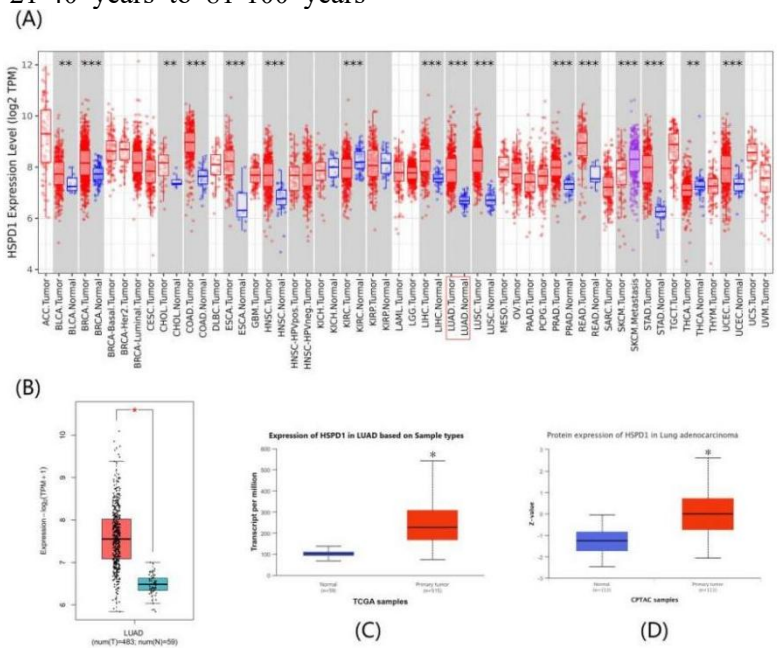

Figure 1. HSPD1 expression in multiple cancers and LUAD. (A) Multi-cancer (TCGA): Tumor > Normal (*P<0.001). (B) LUAD vs. normal (TCGA, n=483/59): Tumor ↑ (*P<0.05). (C) Transcriptome (TCGA): LUAD transcription ↑ (*P<0.05). (D) Proteome (CPTAC): LUAD protein ↑ (*P<0.05).**

Note: Independent samples t-test; *P<0.05, ***P<0.001.

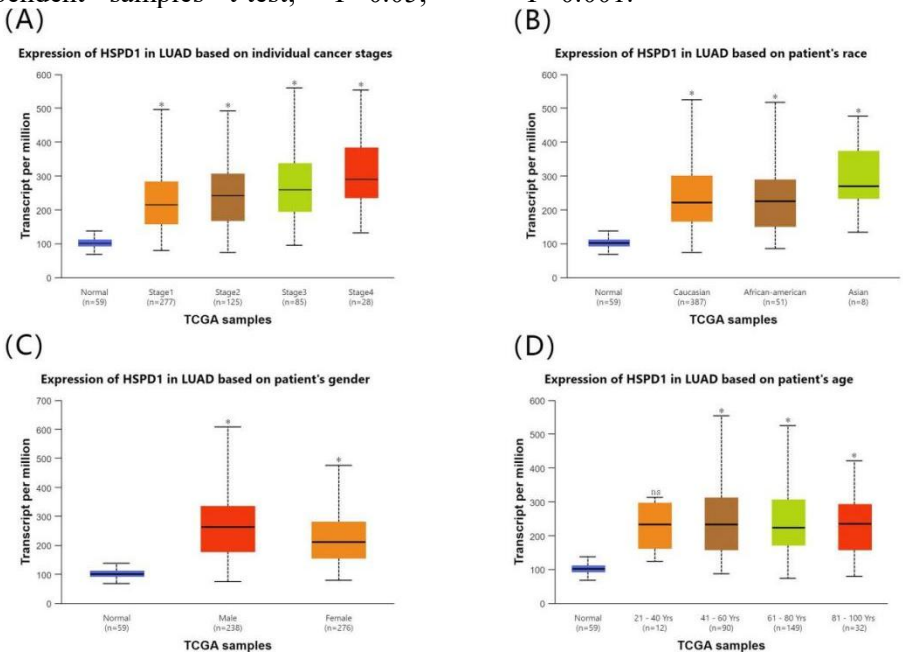

Figure 2. HSPD1 and LUAD clinical features/prognosis. (A) Stage correlation (TCGA): Expression ↑ with stage (*P<0.05). (B) Ethnic comparison (TCGA): Expression ↑ with ethnicities (*P<0.05). (C) Gender comparison (TCGA): Expression ↑ with gender (*P<0.05). (D) Age comparison (TCGA): Expression ↑ with age (*P<0.05).

Note: Log-rank test; P<0.05 significant.

3.3 The Association between HSPD1 Expression and the Survival Prognosis of Lung Adenocarcinoma Patients

To elucidate the correlation between HSPD1 levels and the outlook for LUAD patients, we analyzed survival data from the Kaplan-Meier Plotter database: in terms of Overall Survival (OS, Figure 3A). Patients exhibiting elevated levels of HSPD1 had a notably poorer prognosis compared to those with lower expression, as evidenced by a Hazard Ratio of 1.5 (95% CI: 1.33-1.69), and the logrank test P-value reached 4.9e-11, suggesting that high HSPD1 expression is significantly associated with a shorter patient OS; whereas in First Progression Survival (FPS, Figure 3B), the progression risk of the high HSPD1 expression group was also significantly higher than that of the low expression group, with the corresponding HR of 1.46 (95% CI: 1.23-1.73), and the logrank test P-value was 1.1e-05.

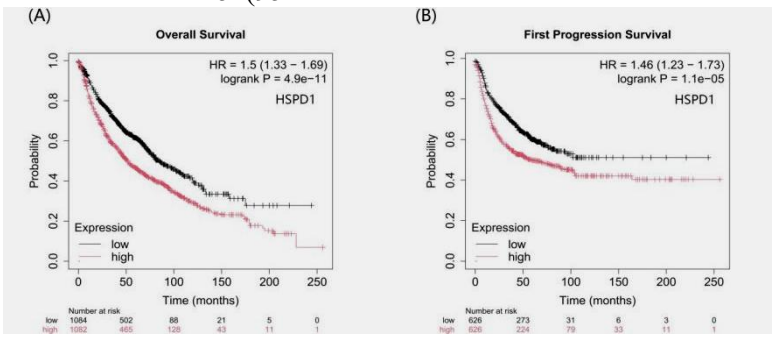

Figure 3. HSPD1 and LUAD Prognosis: (A) OS: HSPD1 High → Shorter OS (HR=1.5, P=4.9e-11). (B) FPS: HSPD1 High → Shorter FPS (HR=1.46, P=1.1e-05).

Note: Log-rank test; HR=hazard ratio.

3.4 Association of HSPD1 Expression with Immune Infiltration and Copy Number Variation in LUAD

Through the TIMER2.0 database, the relationship between HSPD1 and the immunosuppressive microenvironment of LUAD was analyzed (Figure 4 A). The expression level of HSPD1 was significantly negatively correlated with the infiltration levels of B cells, CD4+ T cells, macrophages, and dendritic cells (partial.cor were -0.221, -0.26, -0.155, -0.189, respectively, all P<0.05), but showed no significant association with the infiltration of CD8+ T cells or neutrophils (P>0.05), suggesting that high expression of HSPD1 may be accompanied by a reduction in the infiltration of certain anti-tumor immune cells. Further analysis of the impact of HSPD1 copy number variation on immune infiltration (Figure 4 B) revealed that among different copy number variation subtypes (arm-level deletion, diploid/normal, arm-level amplification, high amplification), the infiltration levels of immune cells such as B cells and CD8+ T cells were significantly lower in samples with HSPD1 copy number amplification (especially high amplification) (P<0.05).

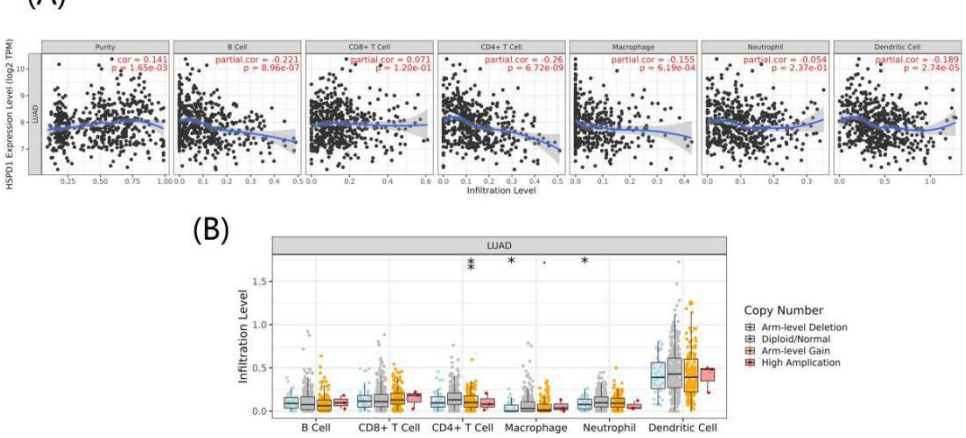

Figure 4. HSPD1 and LUAD Immune Microenvironment. (A) Infiltration Heatmap (TIMER2.0): Negatively Correlated with B/CD4+ T Cells (P<0.05). (B) CNA vs. Infiltration: High Amplification → B/CD8+ T Cells ↓ (*P<0.05).

Note: Pearson test/ANOVA; *P<0.05 significant.

3.5 Molecular Network Characteristics of HSPD1 Interacting Genes

A molecular interaction network for HSPD1 was constructed using the GeneMANIA database (Figure 5), identifying a total of 20 genes with potential interactions with HSPD1. Among them, HSPE1 exhibited the strongest association with HSPD1. Concurrently, the CCT family genes (such as CCT8L2, CCT4, CCT2), BAX, and TOMM40 also formed complex interaction relationships with HSPD1.

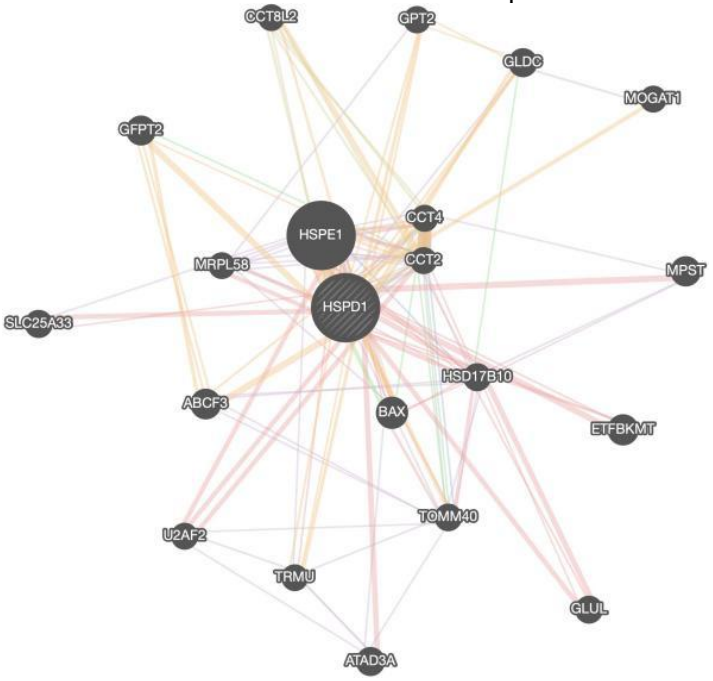

Figure 5. Molecular Network Characteristics of HSPD1 Interacting Genes

3.6 The Functional and Pathway Enrichment Characteristics of HSPD1 Interacting Genes

A GO functional enrichment analysis of the 20 HSPD1 interacting genes (left diagram) shows: at the Biological Process (BP) level, these genes are primarily enriched in entries such as amino acid metabolic processes and protein folding; at the Cellular Component (CC) level, they are concentrated in mitochondrial-related structures (such as mitochondrial matrix, inner membrane); at the Molecular Function (MF) level, they are mainly protein folding molecular chaperones, ATP hydrolysis activity, and other functions, suggesting that HSPD1 interacting genes are mainly involved in cellular functions related to metabolic regulation and protein processing (Figure 6). Further analysis of the KEGG pathway revealed that the genes in question are notably overrepresented in pathways involving amino acid metabolism, like alanine-aspartate-glutamate and arginine synthesis (Figure 7), suggesting that HSPD1 may participate in the biological processes of lung adenocarcinoma by regulating amino acid metabolic pathways.

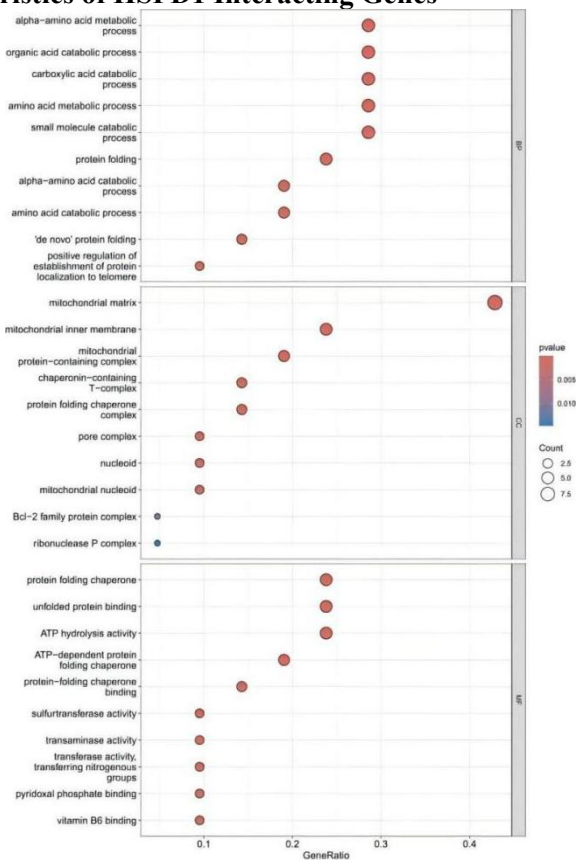

Figure 6. GO Enrichment

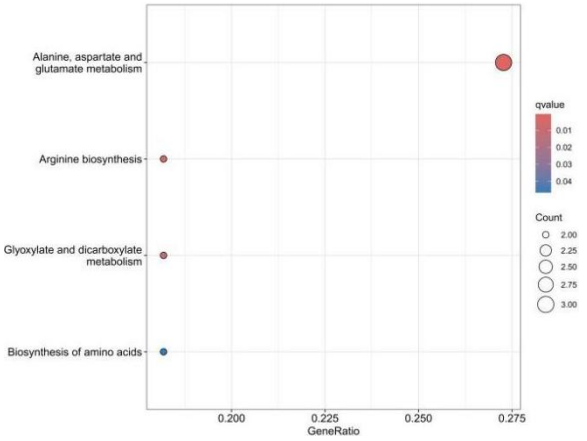

Figure 7. KEGG Enrichment

4. Discussion

Lung adenocarcinoma (LUAD), as the predominant subtype of non-small cell lung cancer, accounts for approximately 40%-50% of the total incidence of lung cancer. Its onset is insidious and progression is rapid, with most patients diagnosed at an advanced stage, leading to a poor prognosis [16]. While cutting-edge treatments like targeted therapy and immunotherapy have turned the tide for numerous patients in recent years, dramatically boosting their chances of survival, issues like tumor heterogeneity and treatment resistance still severely constrain clinical efficacy. There is an urgent need to explore prognostic biomarkers and potential therapeutic targets with clinical value to provide new directions for the precise diagnosis and treatment of LUAD[17]. Heat shock proteins (HSPs), as core molecular chaperones in cellular stress responses, their abnormal expression is closely related to the occurrence and development of tumors [18]. HSPD1 (also known as HSP60), as a key member of this family, has been proven to be highly expressed in various malignant tumors such as colorectal cancer and liver cancer, and is associated with poor prognosis [19]. However, its expression pattern, clinical significance, and regulatory mechanisms in LUAD have not yet been systematically elucidated.

This study systematically investigated the role of HSPD1 in LUAD by integrating multi-omics data from public databases such as TCGA and GEPIA using bioinformatics methods. The results showed that HSPD1 was significantly overexpressed in LUAD and multiple tumors, with consistent validation at both transcriptomic and proteomic levels, which aligns with its expression characteristics in other solid tumors [20], suggesting that its abnormal overexpression may be a common molecular event in tumorigenesis and progression. Clinical feature analysis indicated that HSPD1 expression progressively increased with the advancement of LUAD staging and remained high in patients of different races, genders, and ages, implying its potential involvement in tumor malignant progression and its broad clinical applicability. Survival analysis further confirmed that high HSPD1 expression was significantly associated with shortened overall survival and increased risk of disease progression in LUAD patients, consistent with the conclusions of prognostic value studies on HSPD1 in non-small cell lung cancer, providing an important reference for clinical prognostic stratification.

The remodeling of the tumor immune microenvironment is critical for tumor progression [21]. This study found that HSPD1 expression is significantly negatively correlated with the infiltration of anti-tumor immune cells such as B cells and CD4+ T cells, and that copy number amplification can further reduce immune cell infiltration, which is consistent with the mechanism by which gene copy number variation regulates the immune microenvironment [9]. It is hypothesized that a regulatory axis of "copy number amplification → HSPD1 upregulation → formation of an immunosuppressive microenvironment" exists. Molecular interaction and enrichment analyses showed that HSPD1 core interacting genes are mainly involved in processes such as amino acid metabolism and protein folding, and are enriched in mitochondrial-related components and amino acid metabolism pathways, suggesting that HSPD1 may support LUAD progression by regulating metabolic reprogramming [20].

This study has certain limitations, as the data all originate from public databases and lack validation from in vitro cell experiments and in vivo animal models. Future research needs to further confirm the pro-cancerous role and mechanism of HSPD1 through functional experiments.

5. Conclusion

In summary, HSPD1 is highly expressed in LUAD, is closely related to tumor stage and prognosis, and may participate in LUAD progression by regulating immune cell

infiltration and amino acid metabolism. It is expected to become a potential prognostic biomarker and therapeutic target for LUAD.

References

[1] Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2020.* CA Cancer J Clin, 2020. 70(1): p. 7-30.

[2] Bade, B.C. and C.S. Dela Cruz, *Lung Cancer 2020: Epidemiology, Etiology, and Prevention.* Clin Chest Med, 2020. 41(1): p. 1-24.

[3] Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy.* Science, 2004. 304(5676): p. 1497-500.

[4] Wang, M., et al., *Microglia-Mediated Neuroinflammation: A Potential Target for the Treatment of Cardiovascular Diseases.* J Inflamm Res, 2022. 15: p. 3083-3094.

[5] Omidi, A., et al., *Heat-shock proteins, oxidative stress, and antioxidants in one-humped camels.* Trop Anim Health Prod, 2023. 56(1): p. 29.

[6] Wang, X., et al., *[Retracted] MicroRNA-454 inhibits the malignant biological behaviours of gastric cancer cells by directly targeting mitogen-activated protein kinase 1.* Oncol Rep, 2022. 47(4).

[7] Grzes, M., et al., *A Driver Never Works Alone-Interplay Networks of Mutant p53, MYC, RAS, and Other Universal Oncogenic Drivers in Human Cancer.* Cancers (Basel), 2020. 12(6).

[8] Dong, S., et al., *MMP28 recruits M2-type tumor-associated macrophages through MAPK/JNK signaling pathway-dependent cytokine secretion to promote the malignant progression of pancreatic cancer.* J Exp Clin Cancer Res, 2025. 44(1): p. 60.

[9] Li, F., et al., *Relationship Between CNVs and Immune Cells Infiltration in Gastric Tumor Microenvironment.* Front Genet, 2022. 13: p. 869967.

[10] Tomczak, K., P. Czerwińska, and M. Wiznerowicz, *The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge.* Contemp Oncol (Pozn), 2015. 19(1a): p. A68-77.

[11] Li, T., et al., *TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells.* Cancer Res, 2017. 77(21): p. e108-e110.

[12] Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.* Nucleic Acids Res, 2017. 45(W1): p. W98-w102.

[13] Chandrashekar, D.S., et al., *UALCAN: An update to the integrated cancer data analysis platform.* Neoplasia, 2022. 25: p. 18-27.

[14] Gomes, A.P., et al., *Kaplan-Meier Survival Analysis: Practical Insights for Clinicians.* Acta Med Port, 2024. 37(4): p. 280-285.

[15] Franz, M., et al., *GeneMANIA update 2018.* Nucleic Acids Res, 2018. 46(W1): p. W60-w64.

[16] Luo, G., et al., *Estimated worldwide variation and trends in incidence of lung cancer by histological subtype in 2022 and over time: a population-based study.* Lancet Respir Med, 2025. 13(4): p. 348-363.

[17] *Comprehensive molecular profiling of lung adenocarcinoma.* Nature, 2014. 511(7511): p. 543-50.

[18] Ciocca, D.R. and S.K. Calderwood, *Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications.* Cell Stress Chaperones, 2005. 10(2): p. 86-103.

[19] Yang, W., et al., *Co-expression Network Analysis Identified Key Proteins in Association With Hepatic Metastatic Colorectal Cancer.* Proteomics Clin Appl, 2019. 13(6): p. e1900017.

[20] Parma, B., et al., *Metabolic impairment of non-small cell lung cancers by mitochondrial HSPD1 targeting.* J Exp Clin Cancer Res, 2021. 40(1): p. 248.

[21] Lv, B., et al., *Immunotherapy: Reshape the Tumor Immune Microenvironment.* Front Immunol, 2022. 13: p. 844142.

