Comprehensive analysis to identify a novel diagnostic marker of lung adenocarcinoma and its immune infiltration landscape

Ankang Zhu[1,2†], Dongchen Pei[1†], Yan Zong[1†], Yan Fan[1,2], Shuai Wei[1,2], Zhisong Xing[1,2], Shuailin Song[1,2], Xin Wang[1,2] and Xingcai Gao[1,2*]

[1]The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou University, Zhengzhou, China, [2]Department of Cardiothoracic Surgery, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou University, Zhengzhou, China

Background: Lung cancer continues to be a problem faced by all of humanity. It is the cancer with the highest morbidity and mortality in the world, and the most common histological type of lung cancer is lung adenocarcinoma (LUAD), accounting for about 40% of lung malignant tumors. This study was conducted to discuss and explore the immune-related biomarkers and pathways during the development and progression of LUAD and their relationship with immunocyte infiltration.

Methods: The cohorts of data used in this study were downloaded from the Gene Expression Complex (GEO) database and the Cancer Genome Atlas Program (TCGA) database. Through the analysis of differential expression analysis, weighted gene co-expression network analysis (WGCNA), and least absolute shrinkage and selection operator(LASSO), selecting the module with the highest correlation with LUAD progression, and then the HUB gene was further determined. The Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) were then used to study the function of these genes. Single-sample GSEA (ssGSEA) analysis was used to investigate the penetration of 28 immunocytes and their relationship with HUB genes. Finally, the receiver operating characteristic curve (ROC) was used to evaluate these HUB genes accurately to diagnose LUAD. In addition, additional cohorts were used for external validation. Based on the TCGA database, the effect of the HUB genes on the prognosis of LUAD patients was assessed using the Kaplan-Meier curve. The mRNA levels of some HUB genes in cancer cells and normal cells were analyzed by reverse transcription-quantitative polymerase chain reaction (RT-qPCR).

Results: The turquoise module with the highest correlation with LUAD was identified among the seven modules obtained with WGCNA. Three hundred fifty-four differential genes were chosen. After LASSO analysis, 12 HUB genes were chosen as candidate biomarkers for LUAD expression. According to the immune infiltration results, CD4 + T cells, B cells, and NK cells were high in LUAD sample

tissue. The ROC curve showed that all 12 HUB genes had a high diagnostic value. Finally, the functional enrichment analysis suggested that the HUB gene is mainly related to inflammatory and immune responses. According to the RT-qPCR study, we found that the expression of DPYSL2, OCIAD2, and FABP4 in A549 was higher than BEAS-2B. The expression content of DPYSL2 was lower in H1299 than in BEAS-2B. However, the expression difference of FABP4 and OCIAD2 genes in H1299 lung cancer cells was insignificant, but both showed a trend of increase.

Conclusions: The mechanism of LUAD pathogenesis and progression is closely linked to T cells, B cells, and monocytes. 12 HUB genes(ADAMTS8, CD36, DPYSL2, FABP4, FGFR4, HBA2, OCIAD2, PARP1, PLEKHH2, STX11, TCF21, TNNC1) may participate in the progression of LUAD *via* immune-related signaling pathways.

KEYWORDS

lung adenocarcinoma, immune cell infiltration, biomarkers, immune-related pathways, LASSO, RT-qPCR

1 Introduction

Lung cancer is one of the most common cancers in the world. In recent years, the number of new cases has reached a peak, and lung cancer has the highest number of deaths (1, 2). Lung adenocarcinoma has the highest incidence of lung cancer at approximately 40% (3). While one of the reasons for the high mortality of LUAD is that 57% of cases had progressed at the time of testing, when the treatment regimen was limited, with 1-and 5-year survival rates of only 26% and 4%, respectively (4, 5). This result is not satisfactory, and although the rapid development of immunotherapy and targeted therapies in recent years has led to a significant improvement in the outcomes of LUAD patients, the prognosis outcome of LUAD patients is still unsatisfactory (6, 7). Thus, there is a need to investigate and discover novel biomarkers or immune cell infiltration during LUAD progression, which is of extraordinary importance for the early detection, diagnosis, treatment, and better prognosis of LUAD. Despite the diverse pathogenesis and causes of LUAD, extensive clinical evidence and experimental data show that immunocytes and immune-related pathways play various roles in the development of LUAD and the prognostic process. For example, a reduction in CD4 + T cells suppresses the activity of cytotoxic T cells in tumors, thereby

restricting LUAD tumor cell growth (6). Programmed cell death 1 (PD-1) is expressed in T cells to suppress peripheral autoimmunity (immune tolerance) (8). M2-polarized macrophages exhibit immunosuppressive activity and promote tumor angiogenesis in LUAD patients (9). Many other molecules are closely associated with LUAD and play an immunological role in tumor progression. Thus, further investigation into the molecular mechanisms of LUAD pathogenesis is still needed.

WGCNA works by analyzing a large number of genes and then putting genes with similar expressions into the same module according to the clustering principle. The most significant advantage of this method over simple cluster analysis is that it is biologically meaningful and allows for effective preliminary screening of genes related to target features (10, 11). In many cases, LASSO algorithms are used to describe the degree of correlation between two related variables. The advantage of this algorithm over the traditional Cox regression and logistic regression lies in its ability to reduce the dimension. Both WGCNA and LASSO regression analysis are commonly used for bioinformatics technology analysis. Moreover, the LASSO analysis of the WGCNA genes can make us more accurate in screening the target feature-related genes (12). In the first step, we screened differentially expressed genes and identified key biomarkers for LUAD progression. Based on the results of the Gene Ontology (GO) of differentially expressed genes and the Kyoto Encyclopedia of Genes and Genomes (KEGG), we found that these DEGs mainly focus on some immune processes and immune pathways related to LUAD. We then used ssGSEA analysis to assess the infiltration of immunocytes in the immune environment in the hope of gaining a clearer understanding of the mechanisms of LUAD progression, and the results may provide a way to understand the pathogenesis of LUAD and find new therapeutic targets.

Abbreviations: LUAD, lung adenocarcinoma; DEGs, Differentially expressed genes; WGCNA,weighted gene co-expression network analysis; GEO, Gene-Expression Omnibus; TCGA, The Cancer Genome Atlas Program; LASSO, Least absolute shrinkage and selection operator; AUC, Area under the curve; ssGSEA, single-sample gene-set enrichment analysis algorithm; GO, Gene Ontology; Tregs, regulatory T cell; PPAR, Peroxisome proliferator-activated receptors; FOXP3, Forkhead box P3; KEGG, Kyoto Encyclopedia of Genes and Genomes; ROC, receiver operating characteristic curve; GSEA, Gene Set Enrichment Analysis.

2 Materials and methods

2.1 Data collection

Microarray expression data and clinical information for LUAD were obtained from the GEO and TCGA databases. There were two cohorts in the treatment group, GSE63459 and GSE176348, with 89 specimens (including 45 tumor samples and 44 normal samples). In addition, external validation using the TCGA-LUAD cohort with 598 samples (including 539 tumor samples and 59 normal samples) was performed. All sequencing information for normal samples comes from adjacent tissues.

2.2 Selection of the DEGs

We used the data normalization and probe annotation from the R software (version 4.2.1) "limma" and "GEOquery" packages for the data of GSE63459 and GSE176348, with P-value < 0.05 and |log fold change (FC) | > 1 for the DEGS screening criteria (13, 14).

2.3 Construction of gene co-expression network

We used the WGCNA to process expression profile data from GSE63459 and GSE176348 datasets to establish a weighted co-expression network. Then we investigated the genes that deviate from the top 25% of the median (10). The data integrity is checked by the 'Good SampleGenes' function. We chose a suitable soft threshold value (β) and validated the ability of the soft threshold value. The matrix data was transformed into an adjacency matrix by us, followed by clustering to identify modules based on the topological overlap. Then, the module feature element (ME) is calculated, the similarly expressed modules are combined into the cluster tree according to the ME, and we draw the hierarchical cluster tree graph. Then, the module and phenotype data are combined, and then the gene significant (GS) and module significant (MS) are calculated; the calculation results are used to evaluate whether the gene and clinical information are essential and to analyze the correlation between the module and the model. In addition, We calculated the module membership (MM) for each gene to analyze the GS values of each module.

2.4 Selection and validation of the HUB genes

The gene with the highest inter-module connectivity was selected as the candidate HUB gene. The GS values for biologically significant genes are also generally higher. Therefore, we chose candidate genes with an absolute GS value> 0.20 and an MM value > 0.80. We then intersected these genes with DEGS using the "glmnet" package in the R software package and used the LASSO analysis to determine the final HUB genes (11). We used box plots to probe the HUB gene expression levels in LUAD samples and healthy samples. With the help of 'MedCalc'

software (version 2.0.1), we draw the receiver operating characteristic curves (ROC) to determine these HUB genes' diagnostic specificity and accuracy. A dataset (TCGA-LUAD) is also available for external verification of the HUB gene's expression level and diagnostic value.

2.5 Prognostic analysis

With the help of the "Survival" and "SurvMiner" packages in the "R" software, we divided the samples in TCGA-LUAD into two groups (high and low expression groups) using the median expression of the HUB gene. Lastly, survival curves for HUB and LUAD genes were plotted using the Kaplan-Meier method with the aid of the software package "ggplot2".

2.6 Immunohistochemical staining was performed

Results of immunohistochemical staining of the HUB gene in normal lung tissue and lung cancer specimens from The Human Protein Atlas(www.proteinatlas.org).

2.7 Assessment of immunocytes infiltration and its association with HUB genes

We quantified the infiltration of 28 immunocytes in the GSE63459 and GSE176348 datasets using the ssGSEA algorithm (15). The box plots we established indicate the differences in the expression levels of these immune cells. We also calculated the Spearman correlation of these immune-infiltrating cells with the candidate HUB genes and visualized the calculated results with the 'ggplot2' program package.

2.8 Functional enrichment analysis

We performed GO analysis of DEGs, KEGG analysis, and GSEA analysis through the 'clusterProfiler' and 'enrichplot' package of the R software package (16). We used the immunological signature genomes from the Molecular Signature Database (MsigDB) as the reference, and the significantly enriched genomes had to meet the P <0.05 and the false discovery rate (FDR) q-value <0.05.

2.9 Experimental validation

Several HUB genes (DPYSL2, OCIAD2, and FABP4) were selected for study to verify the HUB genes' role further. Normal human lung epithelial cells (BEAS-2B) and lung adenocarcinoma cells (A549 and H1299) were collected for culture, moreover, extracted RNA from the three cells using the Trizol reagent. For cDNA synthesis, the synthesis was performed using the reverse transcription reagent VAZYME R232. The final PCR reaction was performed on a quantitative real-time PCR instrument. The reaction parameters included the denaturation process
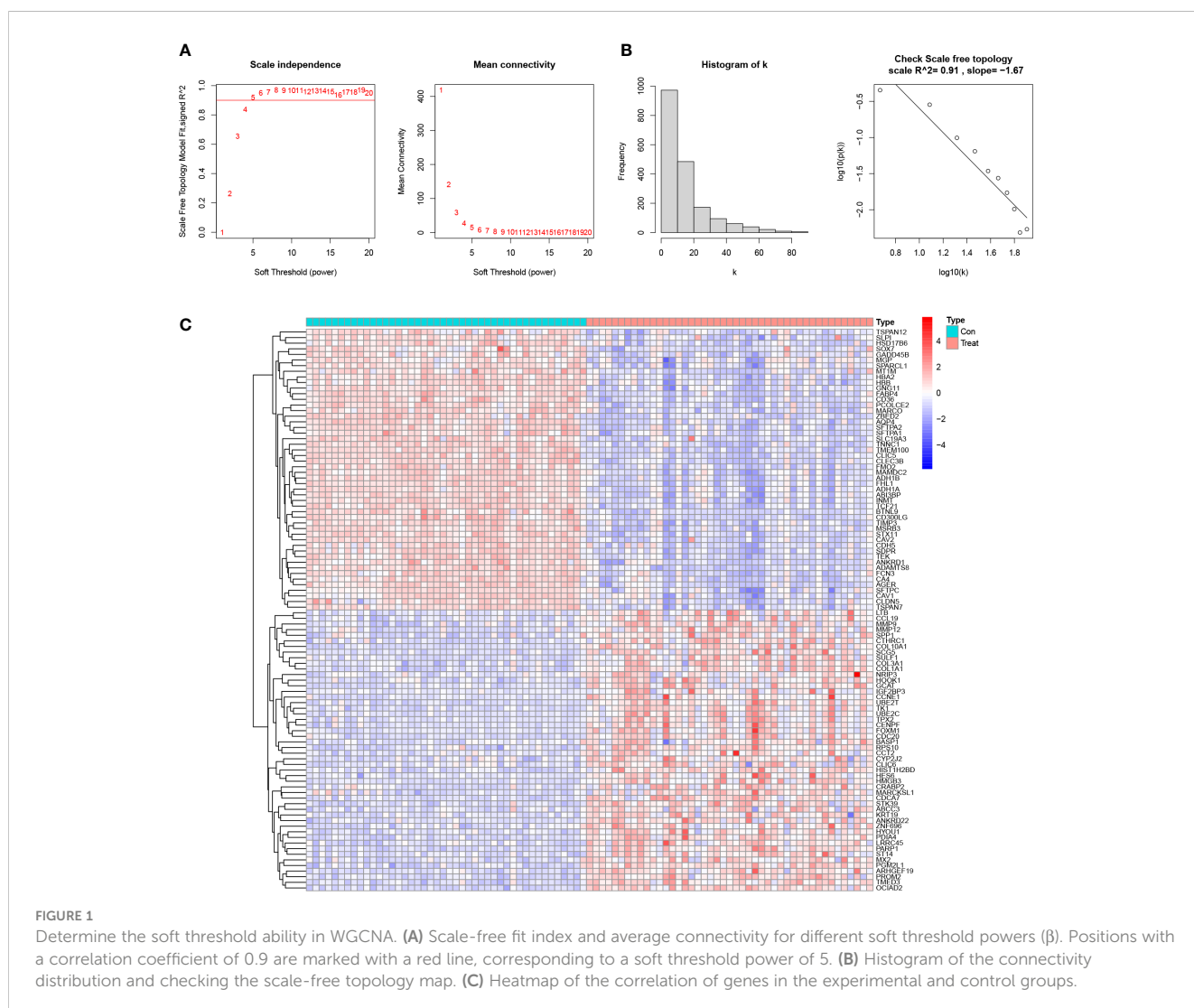
(30s at 95℃), followed by 40 PCR cycles (5s at 95℃ and 34s at 60℃). The melting curve of the PCR product was established, and after the amplification reaction, the temperature was slowly heated from 60℃ at (95℃, 15s, 60℃, 60s, 95℃, 15 s) to 99℃ (instrument automatic ramp rate 0.05℃/s). Quantitative real-time PCR reactions were performed for target and housekeeping genes for each sample. We calculated the relative expression levels of the three genes using the $2 \wedge (-\delta\delta$ ct) method. Since the experimental results of FABP 4 and DPYSL2 were fit to a normal distribution, the analysis was performed using the one-way ANOVA test. We used the Kruskal-Wallis test for statistical analysis for OCIAD2 genes whose results do not conform to the normal distribution (Supplementary file 1). The sequence of the primers is as follows:

DPYSL2, 5'-CCCTGCAGAACATCAACGAC-3' (forward) and 5'-GGCATCTGGAAACGAGTGTG-3' (reverse); OCIAD2, 5'-GTCTGCTCGTGGAAACCAAG-3' (forward) and 5'-CAAGAGACCAGCAAGTGCAAC-3' (reverse); FABP4, 5-GATGACAGGAAAGTCAAGAGCAC-3' (forward) and 5'-GACGCCTTTCATGACGCATTC-3' (reverse); and GADPH,5'-TCTGACTTCAACAGCGACACC-3' (forward) and 5'-CTGTTGCTGTAGCCAAATTCGT-3' (reverse).

3 Results

3.1 Establishment of a co-expression network and selection of key modules

The absolute deviations in the median top 25% of genes were selected for constructing WGCNA, and missing values and outliers in the samples were subsequently removed by cluster analysis. To maintain consistency with the scale-free network, we set the soft threshold β to 5 (scale-free R2 = 0.91; slope =-1.67) (Figures 1A, B). We also analyzed the gene expression in the normal and LUAD samples and plotted the results as a heatmap (Figure 1C). We built a co-expression matrix and obtained seven modules with the help of dynamic hybridization shear (Figure 2A). The relationship between these seven modules and the LUAD and healthy samples is shown in the heatmap. The turquoise module has the highest correlation (cor) (cor = 0.89; P=1e-31) (Figures 2B, C). Moreover, after our correlation analysis, we found that in the turquoise module, GS and MM are also well correlated(cor=1.51; P=1.3e-08) (Figure 2D).

FIGURE 1
Determine the soft threshold ability in WGCNA. **(A)** Scale-free fit index and average connectivity for different soft threshold powers (β). Positions with a correlation coefficient of 0.9 are marked with a red line, corresponding to a soft threshold power of 5. **(B)** Histogram of the connectivity distribution and checking the scale-free topology map. **(C)** Heatmap of the correlation of genes in the experimental and control groups.
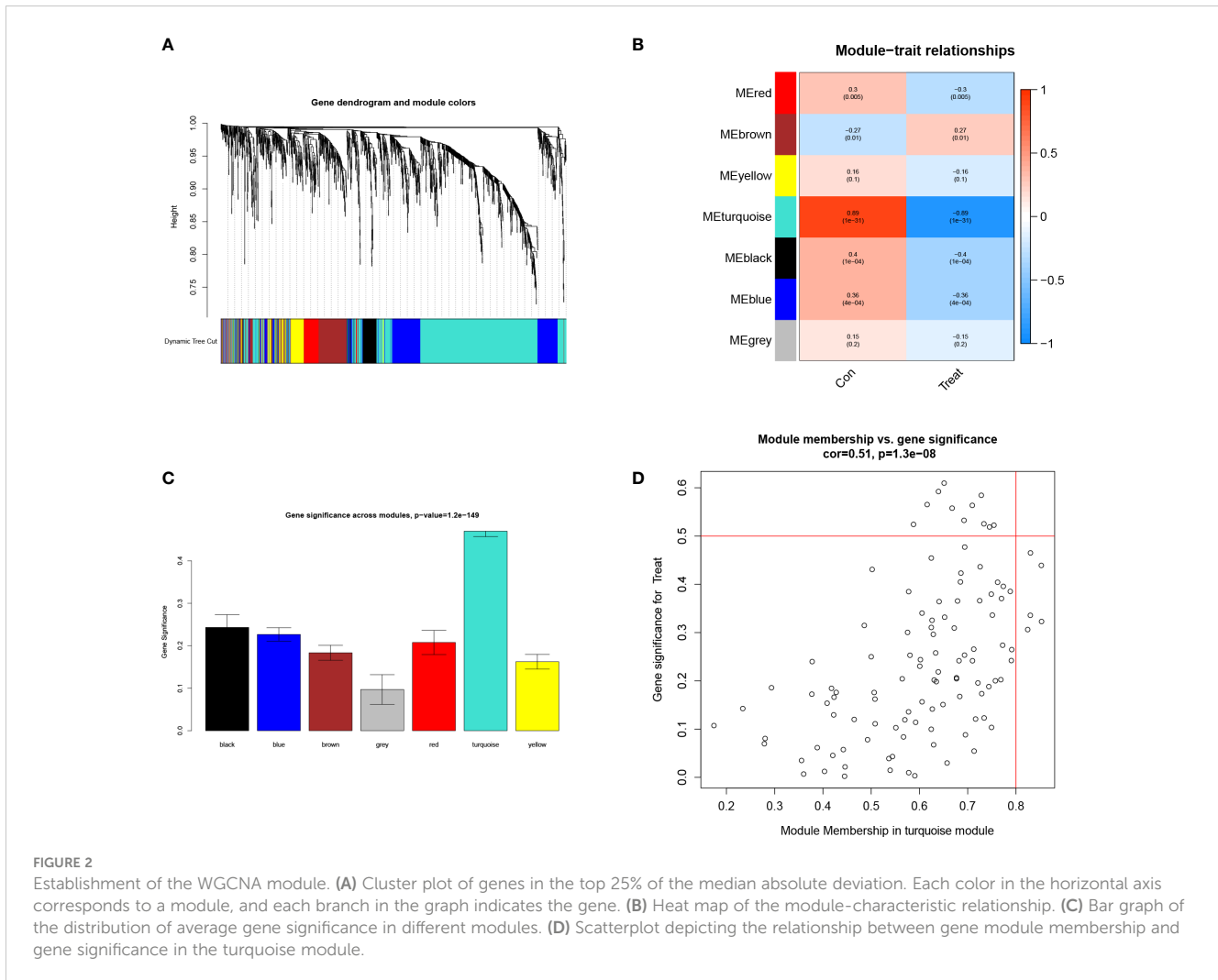
FIGURE 2
Establishment of the WGCNA module. **(A)** Cluster plot of genes in the top 25% of the median absolute deviation. Each color in the horizontal axis corresponds to a module, and each branch in the graph indicates the gene. **(B)** Heat map of the module-characteristic relationship. **(C)** Bar graph of the distribution of average gene significance in different modules. **(D)** Scatterplot depicting the relationship between gene module membership and gene significance in the turquoise module.

3.2 Identification of DEGs and selection of HUB genes

For the DEGs, our filtering criteria were P <0.05 and | logFC |> 1, including 354 differential expressed genes and displaying the results on the volcano plot (Figure 3A). We further selected 87 genes with higher connectivity in the turquoise module using |GS|> 0.20 and |MM|> 0.80 as screening criteria. The results of these two screens were compared, and their intersection was selected as candidate HUB genes, and 85 genes were combined (Figure 3B). Ultimately, after a further screening of these genes using the LASSO analysis, we were able to obtain 12 genes (ADAMTS8, CD36, DPYSL2, FABP4, FGFR4, HBA2, OCIAD2, PARP1, PLEKHH2, STX11, TCF21, TNNC1) (Figures 3C, D).
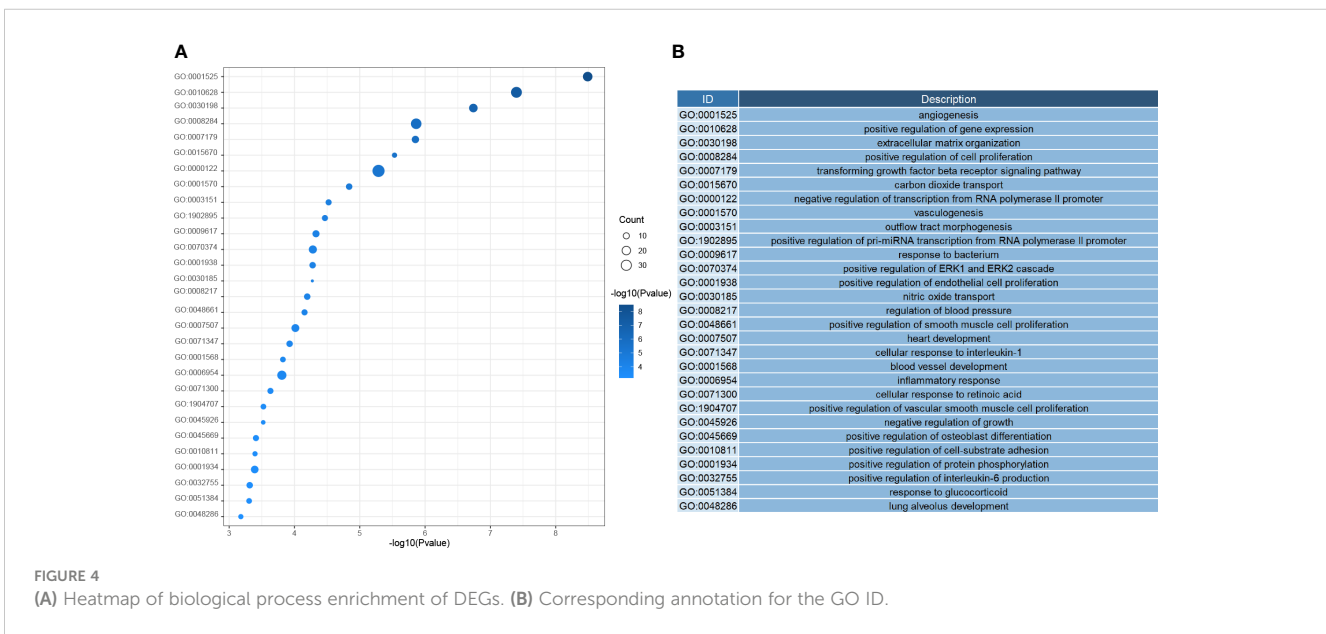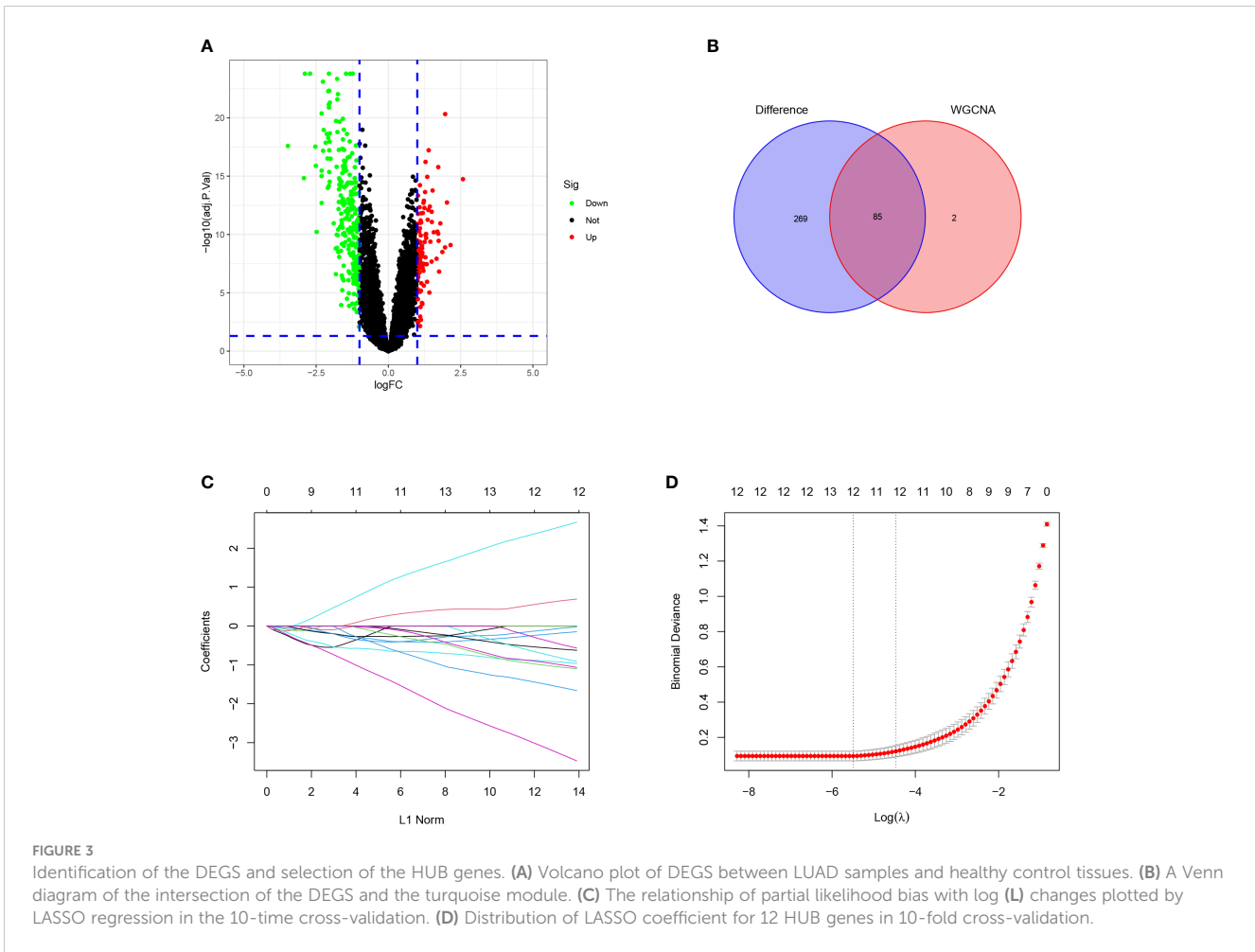
3.3 Functional enrichment analysis of DEGs

Next, we investigated the function of the DEGs screened above; We performed GO analysis on 354 genes. According to the results, we know that DEGs mainly focus on the regulation of genes or pathways (e.g., transforming growth factor receptor signaling pathway, positive regulation of gene expression, negative regulation of transcription by

RNA polymerase II promoter), vascular development (e. g., angiogenesis, angiogenesis, vascular development), immune response and inflammatory response (e. g. inflammation, cellular response of interleukin-1, positive regulation of interleukin-6 production), and even play an essential role in alveolar development (Figure 4). According to the KEGG analysis, We can also learn that DEGs are mainly enriched in the following pathways, pathways of immunological and inflammation-related diseases (AGE-RAGE signaling pathway in diabetic complications, fluid shear stress, and atherosclerosis, rheumatoid arthritis), there are also some immune-related pathways (IL-17 signaling pathway, TNF signal channel, the TGF signaling pathway, PPAR signaling pathway, and other pathways)(Figure 5). The GO analysis and KEGG analysis showed that there are many biological processes and signaling pathways related to the immune-inflammatory response in the development and development of LUAD.

3.4 Immunohistochemical staining of HUB genes in normal tissues and tumor tissues

IHC staining results were paired as shown in Figure 6. On the left of each pair of images is the gene staining in normal lung tissue, and
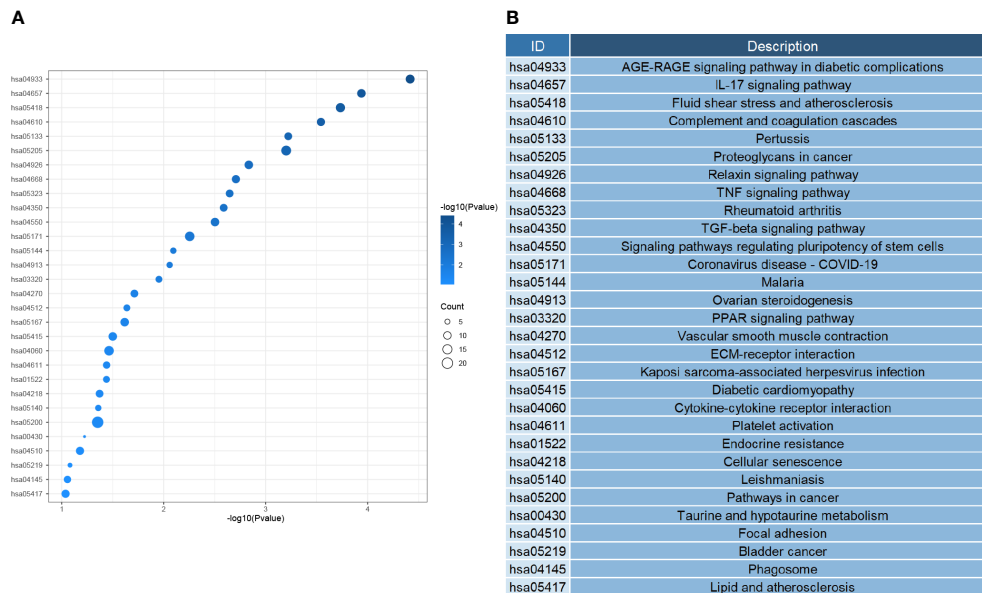
FIGURE 3

Identification of the DEGS and selection of the HUB genes. **(A)** Volcano plot of DEGS between LUAD samples and healthy control tissues. **(B)** A Venn diagram of the intersection of the DEGS and the turquoise module. **(C)** The relationship of partial likelihood bias with log **(L)** changes plotted by LASSO regression in the 10-time cross-validation. **(D)** Distribution of LASSO coefficient for 12 HUB genes in 10-fold cross-validation.

FIGURE 4

(A) Heatmap of biological process enrichment of DEGs. **(B)** Corresponding annotation for the GO ID.

FIGURE 5
(A) Heatmap of signal pathway enrichment of DEGs. **(B)** The KEGG ID corresponds to the annotation.
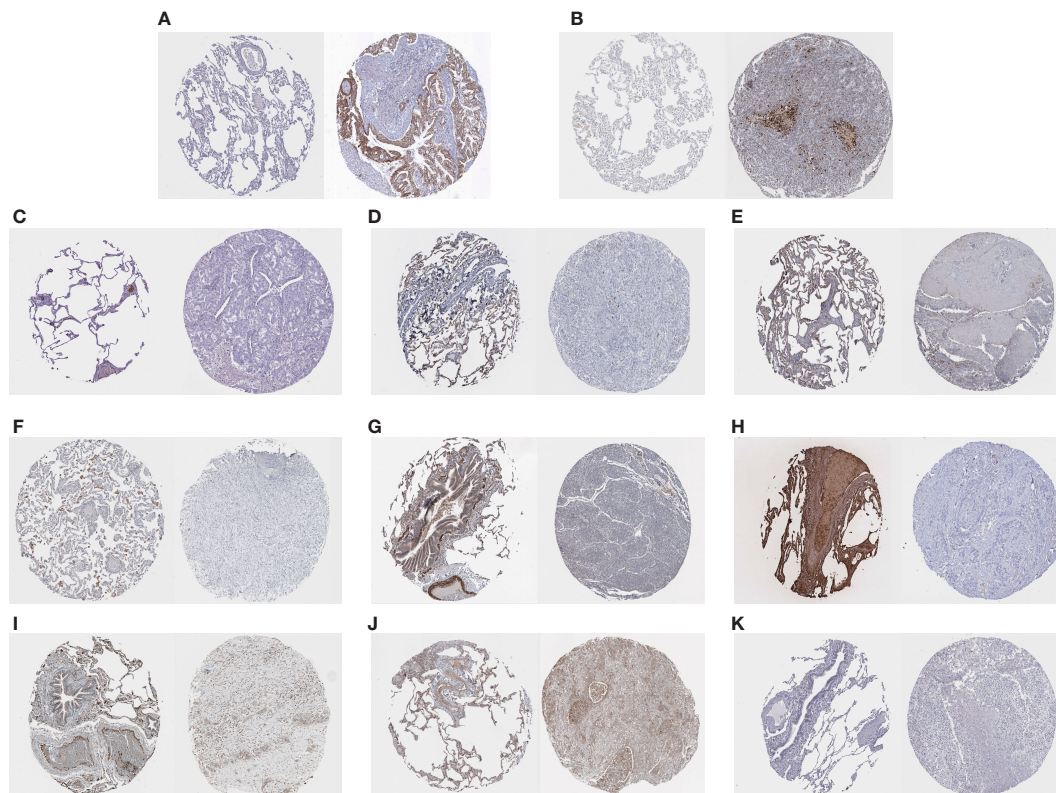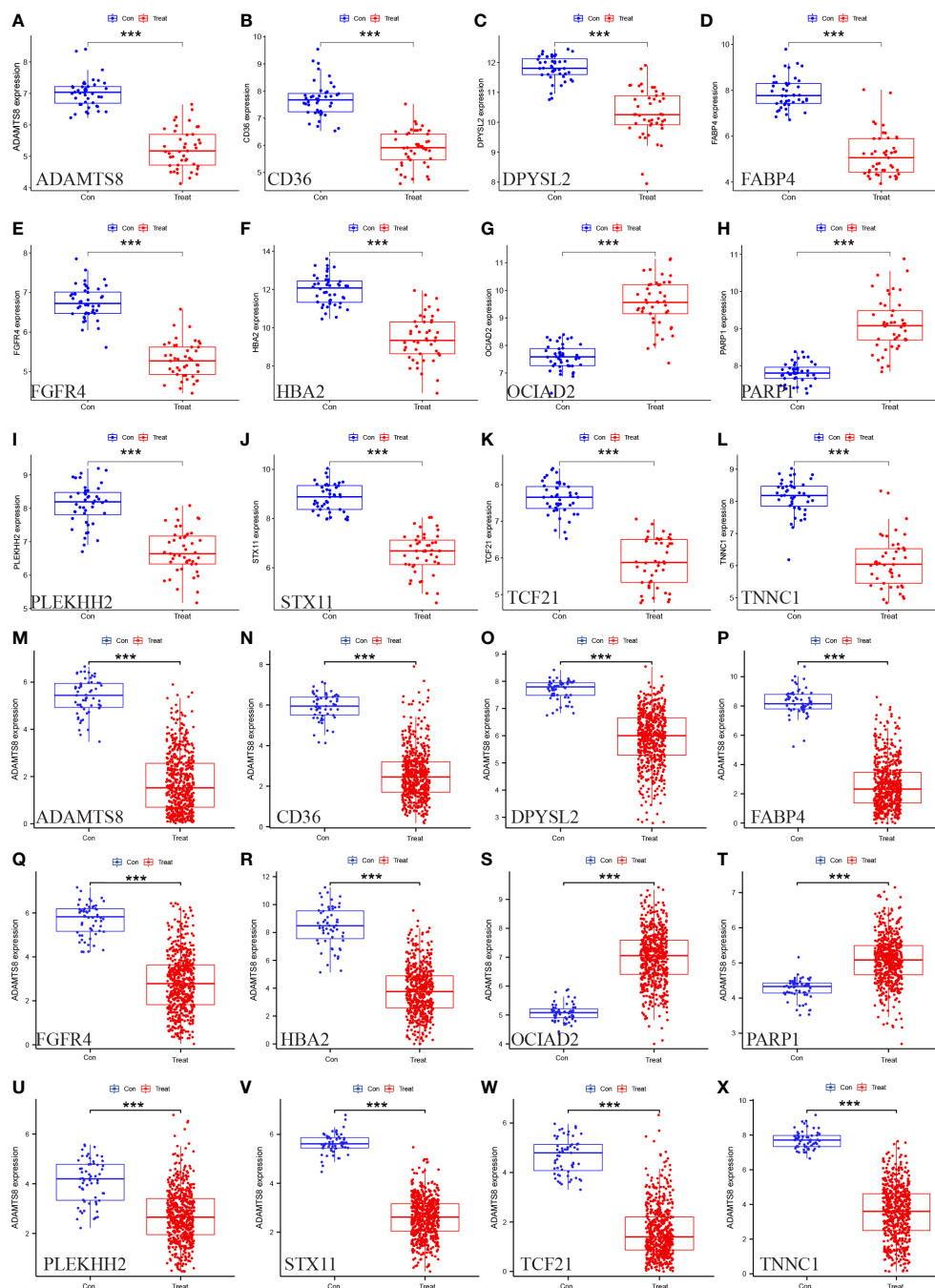

FIGURE 6
(A–K) Results of the immunohistochemical staining of OCIAD2, PARP 1, ADAMTS8, CD36, DPYSL2, FABP4, FGFR4, HBA2, STX11, TCF21, and TNNC 1, on the left of each pair of images, is the staining of genes in normal lung tissue, and on the right is the staining in lung cancer samples.

on the right is the staining in lung cancer samples. We can estimate the expression level of HUB genes, and it is clear that two HUB genes, OCIAD2 and PARP 1, have higher expression in lung cancer samples (Figures 6A, B), while the remaining HUB genes have more expression in normal samples (Figures 6C–K). Unfortunately, we were unable to find the PLEKHH2 staining results, and we will continue to follow up on the results in follow-up studies.

3.5 Validation of HUB gene expression levels and diagnostic value

We assessed the expression levels of the 12 HUB genes by box plots. The results indicated that only OCIAD2, PARP 1 were significantly increased in the control group (Figures 7G, H) ($P < 0.001$), while the other ten genes, ADAMTS8, CD36, DPYSL2,

FIGURE 7
Verification of the 12 HUB genes at the gene expression level. **(A−L)** Verification of the HUB genes in the GSE63459 and GSE176348 **(M−X)** Verification of the HUB gene in the TCGA-LUAD cohort (* represents P <0.05, ** represents P <0.01, and *** represents P <0.001).

FABP4, FGFR4, HBA2, PLEKHH2, STX11, TCF21, TNNC1 were higher in the control group (Figures 7A–F, I–L). Next, we also externally verified the expression levels of these 12 HUB genes using the TCGA-LUAD dataset, and validation results were in agreement with our experimental group (Figures 7M–X). In the ROC curve analysis of the 12 HUB genes, the area under the curve (AUC) of the HUB gene represents the sensitivity and specificity of the HUB gene for the diagnosis of LUAD. From the ROC curve, we can know that the AUC values of all 12 HUB genes were> 0.93, indicating the high value of HUB genes for the diagnosis of LUAD (Figure 8A). While in the TCGA-LUAD cohort, the AUC values, except for PAPR 1 and PLEKHH2, were 0.884 and 0.839, respectively. The AUC values for the remaining HUB genes were all> 0.95, which coincident with the findings from our cohort study above (Figure 8B).

3.6 Prognostic analysis of HUB genes

We partitioned LUAD samples into two groups (high and low expression groups) based on the TCGA-LUAD database. Kaplan-Meier curves were performed for the HUB gene in order to analyze its prognostic relationship to LUAD patients. According to the results, the high expression of OCIAD2 and PARP1 is linked to poor prognosis in LUAD patients (Supplementary Figures 1G, H). However, high expression of ADAMTS8, CD36, DPYSL2, PLEKHH2, STX11, and TCF21 tends to lead to better prognosis (Supplementary Figures 1A–F, I–L), which coincides with the difference in expression of these HUB genes in normal samples and lung cancer samples.

3.7 Immune cell infiltration and its correlation with HUB genes

We compared and analyzed the immune cell infiltration of LUAD samples and the control group with the ssGSEA algorithm. The graph shows the distribution of 28 immunocytes in two datasets, GSE63459 and GSE17634 (Figure 9A). Shown according to its results, the CD4+T cells, CD8+T cells, natural killer (NK) cells, and Bcell in LUAD samples were higher than that in normal samples, and this result indicates that these cells play a significant role in the progression of LUAD (Figure 9B). According to the correlation analysis of HUB genes and immune cell infiltration, most of these HUB genes showed a positive correlation with immune infiltrating cells, such as macrophage, CD4 + T cell, CD8 + T cell, and NK cell. CD8 + T cell exerts antitumor effects in a wide range of cancers. It should be noted that OCIAD2 and PARP1 genes are negatively associated with numerous immune cells. This fits with their results leading to the poor prognosis associated with patients with LUAD (Figure 9C).
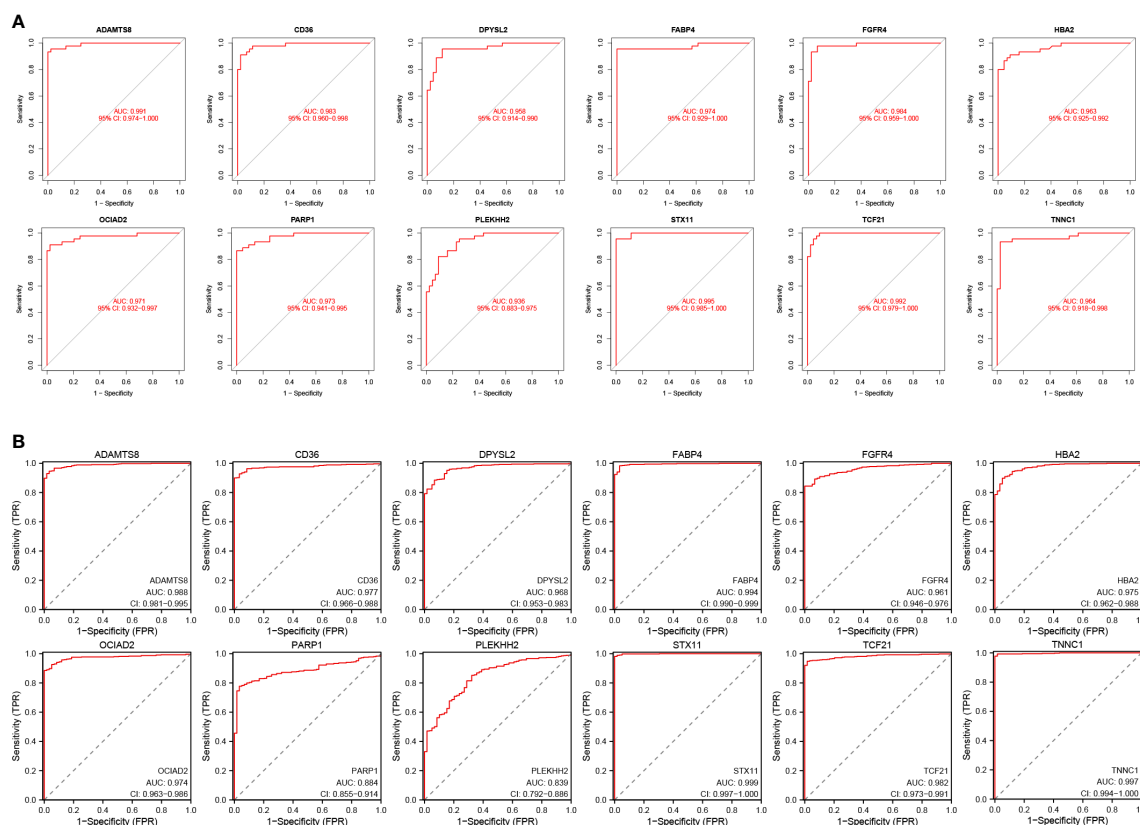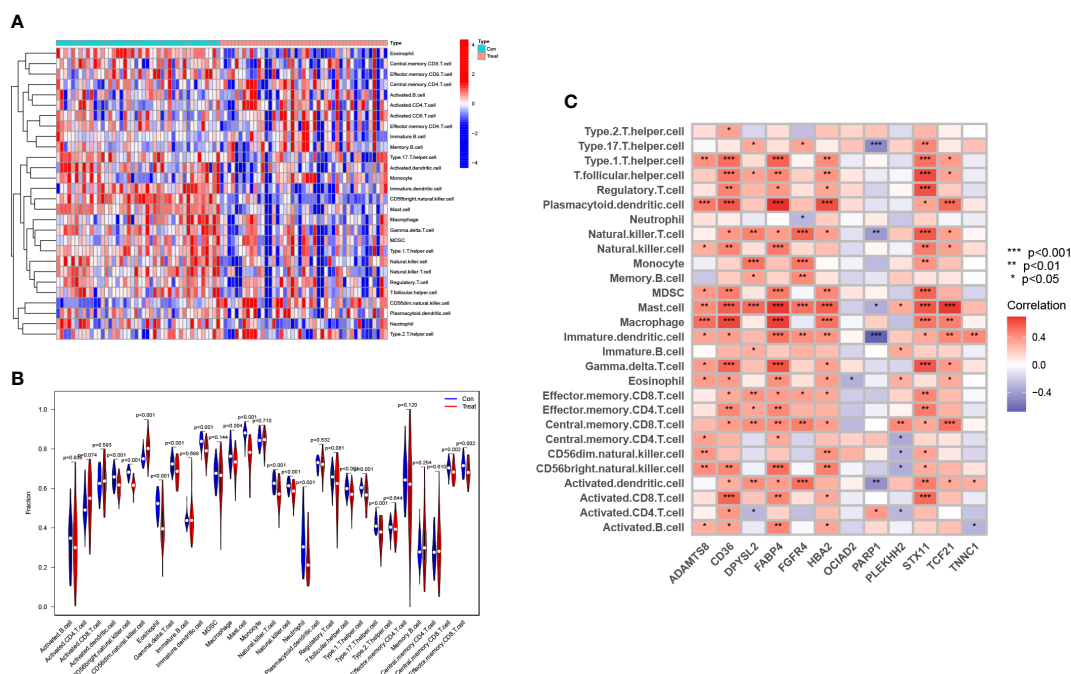

FIGURE 8
Verification of the diagnostic value of the 12 HUB genes. **(A)** Verification of the HUB genes in the GSE63459 and GSE176348 cohorts. **(B)** Validation of the HUB gene in the TCGA-LUAD cohort.
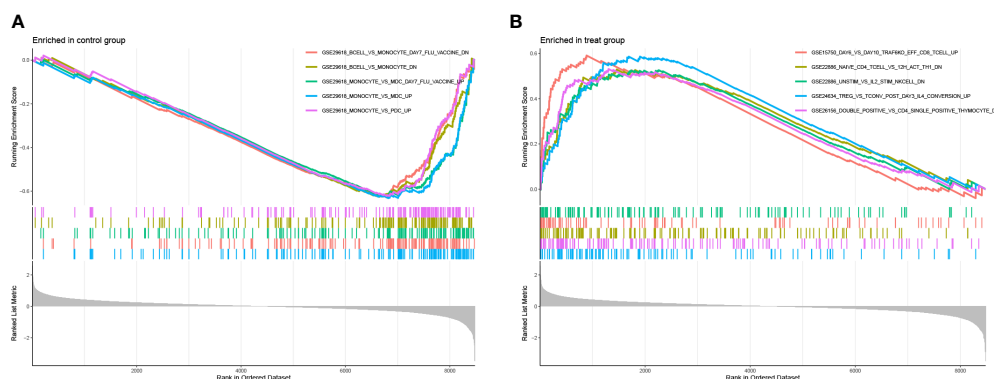
FIGURE 9

Analysis of the immune microenvironment associated with LUAD. Both **(A)** and **(B)** show the distribution of 28 immune cells in the immune microenvironment of normal and LUAD samples. **(C)** The relationship of the 12 HUB genes with the infiltration of multiple immune cells. (* represents P <0.05, ** represents P <0.01, and *** represents P <0.001).

3.8 Enrichment analysis of GSEA immune signature gene sets

To make out that immunogenetics may exist during the progression of LUAD, we have used the immunologically signature gene set in the MsigDB database as a reference standard for DEGS GSEA. A total of 819 gene sets were significantly enriched (|normalized enrichment score (NES)|> 1; FDR Q value <0.05). These genes were mainly concentrated in CD8 T cells, NK Cells, CD4+T cells, monocytes, and regulatory T cells (Supplementary Table S1). Based on the above findings, it appears that many immune genes play an essential role in LUAD progression (Figure 10).

3.9 Detection of mRNA levels of HUB genes by RT-qPCR

After statistical analysis of the PCR results, we found that DPYSL2 (p <0.001), OCIAD2 (p <0.05), and FABP4 (p <0.001) had higher expression in A549 compared to BEAS-2B, showing a statistically significant difference. The expression content of DPYSL2 was lower in H1299 cells compared to t BEAS-2B (p <0.01). Although the expression difference of FABP 4 and OCIAD2 genes in H1299 was not statistically significant, they both showed a trend to increase (Figure 11). These experimental results can better support our study. Nevertheless, the results may require further study with a larger sample size.

FIGURE 10

(A, B) plots represent the enrichment map of the GSEA immune marker database for the experimental and control groups, respectively.
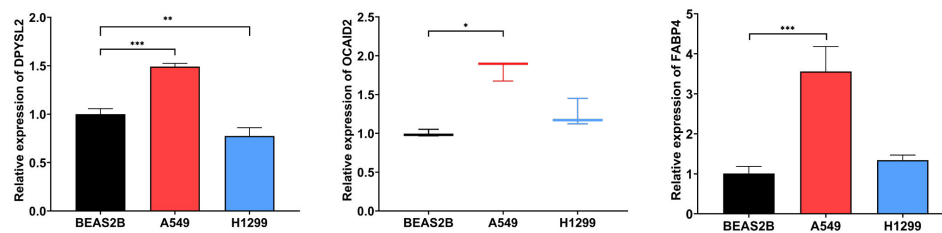
FIGURE 11

The mRNA levels of DPYSL2, OCIAD2, and FABP 4 were measured in human normal lung epithelial cell lines (BEAS-2B) and LUAD cell lines (A549 and H1299), respectively. (* represents P <0.05, ** represents P <0.01, and *** represents P <0.001).

4 Discussion

Using high-throughput microarray technology is a more efficient and accurate bioinformatics method when finding and screening key genes related to the mechanism of cancer development. This technology can also help us diagnose and treat diseases and help us in the design of new drugs. DEG is primarily enriched in leukocyte activation and production, alveolar development, the development of angiogenesis as well as certain immune responses, which are related to the mechanism of LUAD development (17). Analysis of the KEGG showed that DEGs were primarily enriched in immune-related pathways (IL-17 signaling pathway, TNF signal channel, The TGF signaling pathway, PPAR signaling pathway, and other pathways). The cytokine IL-17 can mediate a variety of physiological effects (18, 19). IL-17 is produced primarily by both innate and adaptive immune cells, whose main role is to exert its immune regulatory function by promoting the production of various proinflammatory cytokines and chemokines (IL-6, IL-23, transforming growth factor- β, tumor necrosis factor, etc.), leading to the accumulation of neutrophils and macrophages at the site of inflammation (20–22). IL-17 can stimulate lung tumor growth by promoting angiogenesis and proliferative activation (23, 24). IL-17 in the immune microenvironment can also induce lung cancer metastasis and spread (25). It has also been shown that treatment with a neutralizing anti-IL-17A antibody can reduce the angiogenesis of the tumor as well as reduce the inflammatory response, thereby reducing the growth of lung cancer progression (24, 26). IL-17 is overexpressed in a variety of lung cancer types. During the development of LUAD, the recruitment of numerous neutrophils by IL-17 leads to sustained inflammatory damage (27). All point to a strong link between IL-17 with LUAD progression and prognosis, and these studies are in good agreement with the DEGS enrichment results indicating that there are indeed genes within DEGS that are important in LUAD development.

Traditional DEG-based screening approaches are only capable of local analysis of datasets, which can easily cause the omission and loss of core genes. WGCNA can help us to systematically construct individual biological interaction network maps that can help us to identify core genes that are highly associated with disease prognosis (28, 29). Therefore, we used WGCNA to search for genes highly associated with LUAD and crossed the present results to previous

DEGS to obtain highly related and differential genes. After screening these genes by LASSO analysis, the next 12 HUB genes were identified: ADAMTS8, CD36, DPYSL2, FABP4, FGFR4, HBA2, OCIAD2, PARP1, PLEKHH2, STX11, TCF21, TNNC1. These 12 key genes showed distinct differences in expression levels in LUAD samples and healthy samples. Notably, Of these, only OCIAD2 and PARP1 were found to be significantly highly expressed in tissues from LUAD samples, whereas the remaining 10 genes showed higher levels of expression in the control groups.

ADAMTS8 comes from integrins and metalloproteinases of the thrombospondin motif, and some studies show that ADAMTS8 is closely associated with vascular endothelial growth factor A (VEGFA), and some studies have found that ADAMTS8 expression in lung cancer is very low (30, 31). DPYSL2 is extremely highly associated with breast cancer, which can be expressed in breast cancer cells through axonal guidance. We can also use DPYSL2 to inhibit breast cancer progression and metastasis by inducing reversal. At the same time, numerous studies have demonstrated that phosphorylation of DPYSL2 and DPYSL2 is associated with drug resistance and tumor metastasis (32, 33). OCIAD2 belongs to the ovarian cancer immune response antigen (OCIA) domain family, which consists of 154 amino acids. It fulfills its role in tumor metastasis by promoting STAT3 activation and cell migration. And OCIAD2 is also highly expressed in lung adenocarcinoma but also ovarian mucinous tumors (33, 34). PARP1 is the central enzyme for cellular PAR production and a major target for polyadenosine diphosphate ribosylation during DNA damage. Upon DNA strand breaks, PARP1 performs DNA repair by covalently connecting the ADP-ribose moiety to the acceptor site of some amino acids on itself and other proteins (35, 36). Transcription factor 21 (TCF21) belongs to the class bHLHII superfamily of transcription factors and is expressed in various tissues and organs, it's not only related to different biological processes, such as the development of the reproductive system (support cell differentiation and sex determination), respiratory system, spleen development, it also involved in regulating RNA polymerase to transcription process and so on (37, 38). CD36 is a membrane glycoprotein, as well as a scavenger receptor, which is found in a wide variety of cells. CD36 plays a major role in regulating atherosclerosis *via* a variety of pathways (39, 40).

The above studies we performed showed that DEGS is inextricably linked with inflammatory response, immune response, various cytokines, chemokines, and IL-17 factors. In

this study, the infiltration of 28 immunocytes in the immune microenvironment of LUAD samples was investigated by the ssGSEA algorithm. The results showed that CD4 + T cells, CD4 + T cells, CD8 + T cells, natural killer (NK) cells, and Bcell were more infiltrated in LUAD samples than in normal samples. All of these cells are important in LUAD progression (24, 41, 42). However, following correlation analysis of HUB genes and infiltrating immune cells, in our study, most of these HUB genes were found to positively correlate with immune-infiltrating cells such as Macrophage, CD4 + T cells, CD8 + T cells, and NK cells. While CD8 + T cell has extensive anticancer effects (43). Macrophages play an immune role in a variety of tumors (lung cancer, breast cancer, liver cancer, etc.) (44, 45). Notably, OCIAD2 and PARP 1, which are inversely related to many immune cells, coincide with the results that these two genes are associated with the poor prognosis of lung adenocarcinoma. IL-17 mainly originates from Th17 cells, while Th17 cells mainly originate from CD4 + T cells, and high-level expression of IL-17 leads to inflammatory damage of inflammatory cells like neutrophils and leads to tumor vascular growth, both of which lead to the progression and metastasis of lung tumors. The enrichment of Tregs (regulatory T cell) is correlated with the occurrence, progression, metastasis, and prognosis of various malignancies, including lung cancer (46). Whereas the transcription factor Foxp 3 is the main regulator of Treg cell development and inhibitory activity, and this transcription factor is closely related to autoimmune diseases and a stable immune environment (47). In addition to producing plasma cells involved in the pathological process of LUAD, B cells can produce various cytokines that stimulate Tcell activation, thereby reducing the anti-inflammatory properties of regulatory Tcell and promoting the proliferation and differentiation of effector T cells. The above findings indicate that T cell homeostasis in the immune microenvironment is related to the occurrence, development, and prognosis of LUAD (48). Finally, to investigate the possible immune mechanisms during the development of LUAD, we used the immunological marker gene set in the MsigDB database as a reference for DSGS GSEA and found that activated CD8 T cells, NK Cells, CD4+T cells, monocytes, and regulatory T cells had enhanced expression in DEGS. This suggests that LUAD progression may be linked to the activation of T lymphocytes, monocytes, B lymphocytes, and various cytokines produced by themselves or by their interactions. These studies suggest that these HUB genes may have a potential relationship with the development of LUAD.

To conclude, we selected turquoise by WGCNA and LASSO regression analysis, combined with multiple bioinformatic analyses, and ultimately selected the 12 HUB genes with the highest correlation to LUAD (ADAMTS8, CD36, DPYSL2, FABP4, FGFR4, HBA 2, OCIAD2, PARP1, PLEKHH2, STX11, TCF21, TNNC1), and we analyzed and verified the functions of these genes in the present study. The results of this study will provide initial insights and novel insights into the underlying immunomodulatory mechanisms of LUAD. We will further investigate and explore the more sensitive and specific diagnostic markers of LUAD to provide new directions for LUAD diagnosis, treatment regimen, prognosis, and drug design.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

AZ and DP designed this work. YZ performed the validation of the experiments. YF and SW integrated and analyzed the data. ZX, SS and XW wrote this manuscript. XG edited and revised the manuscript. All authors approved this manuscript.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1199608/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Prognostic relationships of 12 HUB genes and LUAD patients.

SUPPLEMENTARY TABLE 1
Enrichment analysis of GSEA immune signature gene sets.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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):209–49. doi: 10.3322/caac.21660

2. Cao W, Chen H-D, Yu Y-W, Li N, Chen W-Q. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J (Engl)* (2021) 134(7):783–91. doi: 10.1097/CM9.0000000000001474

3. Collins LG, Haines C, Perkel R, Enck RE. Lung cancer: diagnosis and management. *Am Fam Physician* (2007) 75(1):56–63.

4. Rodriguez-Canales J, Parra-Cuentas E, Wistuba II. Diagnosis and molecular classification of lung cancer. *Cancer Treat Res* (2016) 170:25–46. doi: 10.1007/978-3-319-40389-2_2

5. Bade BC, Dela CC. Lung cancer 2020: epidemiology, etiology, and prevention. *Clin Chest Med* (2020) 41(1):1–24. doi: 10.1016/j.ccm.2019.10.001

6. Chen J, Fu Y, Hu J, He J. Hypoxia-related gene signature for predicting LUAD patients' prognosis and immune microenvironment. *Cytokine* (2022) 152:155820. doi: 10.1016/j.cyto.2022.155820

7. Lemjabbar-Alaoui H, Hassan OU, Yang YW, Buchanan P. Lung cancer: biology and treatment options. *Biochim Biophys Acta* (2015) 1856(2):189–210. doi: 10.1016/j.bbcan.2015.08.002

8. Nagano T, Tachihara M, Nishimura Y. Molecular mechanisms and targeted therapies including immunotherapy for non-small cell lung cancer. *Curr Cancer Drug Targets* (2019) 19(8):595–630. doi: 10.2174/1568009619666181210114559

9. Xu F, Cui W-Q, Wei Y, Cui J, Qiu J, Hu L-L, et al. Astragaloside IV inhibits lung cancer progression and metastasis by modulating macrophage polarization through AMPK signaling. *J Exp Clin Cancer Res* (2018) 37(1):207. doi: 10.1186/s13046-018-0878-0

10. Langfelder P, Horvath S. WGCNA: an r package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

11. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model *via* coordinate descent. *J Stat Softw* (2011) 39(5):1–13. doi: 10.18637/jss.v039.i05

12. Duan J, Soussen C, Brie D, Idier J, Wan M, Wang Y-P. Generalized LASSO with under-determined regularization matrices. *Signal Process* (2016) 127:239–46. doi: 10.1016/j.sigpro.2016.03.001

13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007

14. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics* (2007) 23(14):1846–7. doi: 10.1093/bioinformatics/btm254

15. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, Angell H, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* (2013) 39(4):782–95. doi: 10.1016/j.immuni.2013.10.003

16. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2(3):100141. doi: 10.1016/j.xinn.2021.100141

17. Kinoshita T, Goto T. Molecular mechanisms of pulmonary fibrogenesis and its progression to lung cancer: a review. *Int J Mol Sci* (2019) 20(6):20061461. doi: 10.3390/ijms20061461

18. Park H, Li Z, Yang XO, Chang SH, Nurieva R, Wang YH, et al. A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol* (2005) 6(11):1133–41. doi: 10.1038/ni1261

19. Gu C, Wu L, Li X. IL-17 family: cytokines, receptors and signaling. *Cytokine* (2013) 64(2):477–85. doi: 10.1016/j.cyto.2013.07.022

20. Jin W, Dong C. IL-17 cytokines in immunity and inflammation. *Emerg Microbes Infect* (2013) 2(9):e60. doi: 10.1038/emi.2013.58

21. Majumder S, McGeachy MJ. IL-17 in the pathogenesis of disease: good intentions gone awry. *Annu Rev Immunol* (2021) 39:537–56. doi: 10.1146/annurev-immunol-101819-092536

22. Ouyang W, Kolls JK, Zheng Y. The biological functions of T helper 17 cell effector cytokines in inflammation. *Immunity* (2008) 28(4):454–67. doi: 10.1016/j.immuni.2008.03.004

23. Zitvogel L, Kroemer G. Lower airway dysbiosis exacerbates lung cancer. *Cancer Discovery* (2021) 11(2):224–6. doi: 10.1158/2159-8290.CD-20-1641

24. Joerger M, Finn SP, Cuffe S, Byrne AT, Gray SG. The IL-17-Th1/Th17 pathway: an attractive target for lung cancer therapy? *Expert Opin Ther Targets* (2016) 20 (11):1339–56. doi: 10.1080/14728222.2016.1206891

25. Salazar Y, Zheng X, Brunn D, Raifer H, Picard F, Zhang Y, et al. Microenvironmental Th9 and Th17 lymphocytes induce metastatic spreading in lung cancer. *J Clin Invest* (2020) 130(7):3560–75. doi: 10.1172/JCI124037

26. Reppert S, Boross I, Koslowski M, Türeci Ö, Koch S, Lehr HA, et al. A role for T-bet-mediated tumour immune surveillance in anti-IL-17A treatment of lung cancer. *Nat Commun* (2011) 2:600. doi: 10.1038/ncomms1609

27. Lin Q, Xue L, Tian T, Zhang B, Guo L, Lin G, et al. Prognostic value of serum IL-17 and VEGF levels in small cell lung cancer. *Int J Biol Markers* (2015) 30(4):e359–63. doi: 10.5301/jbm.5000148

28. Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* (2009) 10:327. doi: 10.1186/1471-2164-10-327

29. Botía JA, Vandrovcova J, Forabosco P, Guelfi S, D'Sa K, Hardy J, et al. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst Biol* (2017) 11(1):47. doi: 10.1186/s12918-017-0420-6

30. Xia MD, Yu RR, Chen DM. Identification of hub biomarkers and immune-related pathways participating in the progression of antineutrophil cytoplasmic antibody-associated glomerulonephritis. *Front Immunol* (2021) 12:809325. doi: 10.3389/fimmu.2021.809325

31. Tang H, Lee M, Kim EH, Bishop D, Rodgers GM. siRNA-knockdown of ADAMTS-13 modulates endothelial cell angiogenesis. *Microvasc Res* (2017) 113:65–70. doi: 10.1016/j.mvr.2017.05.007

32. Wu Y-J, Nai A-T, He G-C, Xiao F, Li Z-M, Tang S-Y, et al. DPYSL2 as potential diagnostic and prognostic biomarker linked to immune infiltration in lung adenocarcinoma. *World J Surg Oncol* (2021) 19(1):274. doi: 10.1186/s12957-021-02379-z

33. Rmaileh AA, Solaimuthu B, Khatib A, Lavi S, Tanna M, Hayashi A, et al. DPYSL2 interacts with JAK1 to mediate breast cancer cell migration. *J Cell Biol* (2022) 221(7):202106078. doi: 10.1083/jcb.202106078

34. Maki M, JeongMin H, Nakagawa T, Kawai H, Sakamoto N, Sato Y, et al. Aberrant OCIAD2 demethylation in lung adenocarcinoma is associated with outcome. *Pathol Int* (2022) 72(10):496–505. doi: 10.1111/pin.13262

35. Spiegel JO, Van Houten B, Durrant JD. PARP1: structural insights and pharmacological targets for inhibition. *DNA Repair (Amst)* (2021) 103:103125. doi: 10.1016/j.dnarep.2021.103125

36. Alemasova EE, Lavrik OI. Poly(ADP-ribosyl)ation by PARP1: reaction mechanism and regulatory proteins. *Nucleic Acids Res* (2019) 47(8):3811–27. doi: 10.1093/nar/gkz120

37. Molkentin JD, Bugg D, Ghearing N, Dorn LE, Kim P, Sargent MA, et al. Fibroblast-specific genetic manipulation of p38 mitogen-activated protein kinase *In vivo* reveals its central regulatory role in fibrosis. *Circulation* (2017) 136(6):549–61. doi: 10.1161/CIRCULATIONAHA.116.026238

38. Ao X, Ding W, Zhang Y, Ding D, Liu Y. TCF21: a critical transcription factor in health and cancer. *J Mol Med (Berl)* (2020) 98(8):1055–68. doi: 10.1007/s00109-020-01934-7

39. Park YM. CD36, a scavenger receptor implicated in atherosclerosis. *Exp Mol Med* (2014) 46(6):e99. doi: 10.1038/emm.2014.38

40. Wang J, Li Y. CD36 tango in cancer: signaling pathways and functions. *Theranostics* (2019) 9(17):4893–908. doi: 10.7150/thno.36037

41. Roy-Chowdhuri S. Molecular pathology of lung cancer. *Surg Pathol Clin* (2021) 14(3):369–77. doi: 10.1016/j.path.2021.05.002

42. Paver E, O'Toole S, Cheng XM, Mahar A, Cooper WA. Updates in the molecular pathology of non-small cell lung cancer. *Semin Diagn Pathol* (2021) 38(5):54–61. doi: 10.1053/j.semdp.2021.04.001

43. van der Leun AM, Thommen DS, Schumacher TN. CD8(+) T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer* (2020) 20(4):218–32. doi: 10.1038/s41568-019-0235-4

44. Choi J, Gyamfi J, Jang H, Koo JS. The role of tumor-associated macrophage in breast cancer biology. *Histol Histopathol* (2018) 33(2):133–45. doi: 10.14670/HH-11-916

45. Cheng K, Cai N, Zhu J, Yang X, Liang H, Zhang W. Tumor-associated macrophages in liver cancer: from mechanisms to therapy. *Cancer Commun (Lond)* (2022) 42(11):1112–40. doi: 10.1002/cac2.12345

46. Wang X, Xiao Z, Gong J, Liu Z, Zhang M, Zhang Z. A prognostic nomogram for lung adenocarcinoma based on immune-infiltrating treg-related genes: from bench to bedside. *Transl Lung Cancer Res* (2021) 10(1):167–82. doi: 10.21037/tlcr-20-822

47. Deng G, Song X, Fujimoto S, Piccirillo CA, Nagai Y, Greene MI. Foxp3 post-translational modifications and treg suppressive activity. *Front Immunol* (2019) 10:2486. doi: 10.3389/fimmu.2019.02486

48. Lee GR. The balance of Th17 versus treg cells in autoimmunity. *Int J Mol Sci* (2018) 19(3):19030730. doi: 10.3390/ijms19030730