# Identification of VWF as a Novel Biomarker in Lung Adenocarcinoma by Comprehensive Analysis

Yi He [1,2], Ruijie Liu [1,2], Mei Yang [1,2], Wu Bi [1,2], Liuyin Zhou [2], Sai Zhang [1,2], Jin Jin [1,2], Xujun Liang [1,2]* and Pengfei Zhang [1,2]*

[1] NHC Key Laboratory of Cancer Proteomics, Department of Oncology, Xiangya Hospital, Central South University, Changsha, China, [2] National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

Lung adenocarcinoma (LUAD) is one of the most malignant tumors with high morbidity and mortality worldwide due to the lack of reliable methods for early diagnosis and effective treatment. It's imperative to study the mechanism of its development and explore new biomarkers for early detection of LUAD. In this study, the Gene Expression Omnibus (GEO) dataset GSE43458 and The Cancer Genome Atlas (TCGA) were used to explore the differential co-expressed genes between LUAD and normal samples. Three hundred sixty-six co-expressed genes were identified by differential gene expression analysis and Weighted Gene Co-expression Network Analysis (WGCNA) method. Those genes were mainly enriched in ameboidal-type cell migration (biological process), collagen-containing extracellular matrix (cell component), and extracellular matrix structure constituent (molecular function). The protein-protein network (PPI) was constructed and 10 hub genes were identified, including IL6, VWF, CDH5, PECAM1, EDN1, BDNF, CAV1, SPP1, TEK, and SELE. The expression level of hub genes was validated in the GEPIA database, compared with normal tissues, VWF is lowly expressed and SPP1 is upregulated in LUAD tissues. The survival analysis showed increased expression of SPP1 indicated unfavorable prognosis whereas high expression of VWF suggested favorable prognosis in LUAD ($p < 0.05$). Based on the immune infiltration analysis, the relationship between SPP1 and VWF expression and macrophage, neutrophil, and dendritic cell infiltration was weak in LUAD. Quantitative real-time PCR (qRT-PCR) and western blotting were used to validate the expression of VWF and SPP1 in normal human bronchial epithelial (HBE) cell and three LUAD cell lines, H1299, H1975, and A549. Immunohistochemistry (IHC) was further performed to detect the expression of VWF in 10 cases LUAD samples and matched normal tissues. In summary, the data suggest that VWF is a potential novel biomarker for prognosis of LUAD.

Keywords: lung adenocarcinoma, differential expression genes analysis, co-expression analysis, VWF, bioinformatics

## INTRODUCTION

Lung cancer is the leading cause of cancer death around the world (1). Lung adenocarcinoma (LUAD) is the common histological type of lung cancer, and LUAD comprising 40% of all lung cancer cases is the most common type of lung cancer. Despite advances in early diagnosis and treatment methods, the 5-year overall survival rate of LUAD patients remains low (2). Therefore, it's imperative to discover new biomarkers for early diagnosis and treatment to improve the prognosis of patients with lung cancer.

Gene expression profiling based on high-throughput techniques has become a powerful tool to identify significant genes associated with LUAD progression. A study combining The Cancer Genome Atlas (TCGA) dataset and the Gene Expression Omnibus (GEO) dataset identified CX3CL1 overexpression as a positive prognostic factor in patients with LUAD (3). By regulating the cell cycle signaling pathway, KIF2C was identified as involved in the invasion and prognosis in LUAD (4). Recently, PKMYT1 was verified as a promising target to improve the radiosensitivity of LUAD (5). Weighted Gene Co-Expression Network Analysis (WGCNA) is an efficient method to explore the relationship between genes and phenotypes (6). By detecting the interested module of clinical trait and co-expressed module of related genes, WGCNA can help us to mine the key genes in cancer (7, 8) and predict the function of target genes, which identifies potential biomarker genes or therapeutic target (9, 10).

In this study, we performed the integrated analysis by combining gene expression profiling from TCGA and GEO datasets with the WGCNA method to gain differential co-expression genes. Next, Gene Ontology (GO) and Kyoto Encyclopedia of Genes (KEGG) analysis of these genes were performed by clusterProlifer package. After that, protein-protein interaction (PPI) was constructed by STRING, survival analysis was performed with Survival package, immune infiltration analysis was used by TIMER. Then, qRT-PCR and western blotting were used to validate the expression of VWF and SPP1 in HBE cell and three LUAD cell lines. Finally, IHC was further detected the expression of VWF in 10 cases LUAD samples and matched normal tissues. Our results suggest that VWF could be a potential novel biomarker for prognosis of human LUAD. The workflow diagram is shown in **Figure 1**.

## MATERIALS AND METHODS

### Datasets Source and Data Preprocessing

The RNAseq data and the gene expression profile of GSE43458 were downloaded from TCGA and GEO. Each dataset was processed using the R software (version 3.6.3, https://www.r–project.org/). TGCA-LUAD consists of 535 LUAD samples and 59 normal tissues, and RNAseq count data corresponding to 19,600 genes. The edgeR package (11) was used to screen differential expression genes, we selected the genes with a cpm (count per million) ≥1 for further study. Rpkm function was used
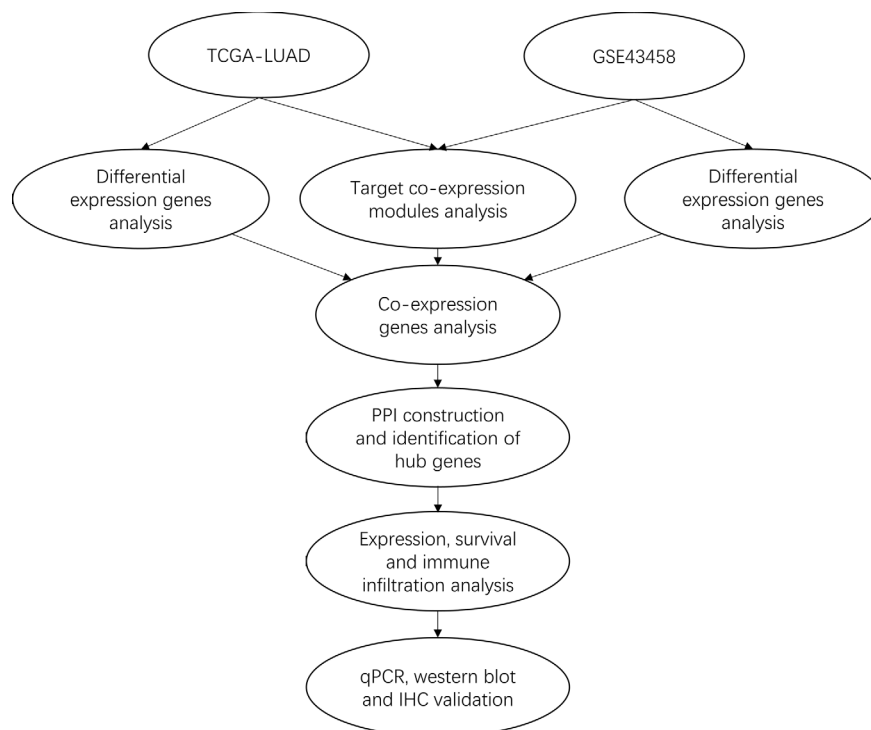


**FIGURE 1** | The workflow of the study.

to filter the data, which is calculated by dividing gene counts by gene length, 15,127 genes with rpkm values were picked for the next study. GSE43458 contains 80 lung adenocarcinoma samples and 30 normal lung tissues. The matrix files of GSE43458 were downloaded by using the GEOquery (12) package and the AnnoProbe package was used to convert the gene probe IDs to the gene symbols. After that, the limma package (13)was used to identify DEGs. As a result, 18,821 genes were selected for further analysis.

## Identification of Key Co-expression Modules Using WGCNA

We constructed gene co-expression networks by the WGCNA package based on the gene expression profile of TGCA-LUAD and GSE43458. The genes with a standard deviation >0.5 were set the criteria. WGCNA was used to explore the relationship between genes and sample traits by grouping highly functionally related genes into modules. Firstly, soft powers β = 6 and 12 were selected using the pick SoftThreshold function to obtain a scale-free network. Next, the adjacency matrix was built according to the Pearson correlation of all gene pairs and transformed into the topological overlap matrix (TOM) as well as the corresponding dissimilarity (1-TOM). Then, performing hierarchical clustering dendrogram on the 1-TOM matrix to divide similar gene expression into different gene co-expression modules. DynamicTreeCut algorithm was used to conduct hierarchical clustering about similar expression genes and classify them into the same gene modules. The trait-modules relationship was estimated by the Pearson correlation coefficient, the module with a high value was selected for subsequent study.

## Screening for Interest Differential Expression Genes

An adjusted p-value <0.05 and the absolute of log2 fold change (log2FC) >1 were set as the cut-off criteria. The DEGs of TCGA-LUAD and GSE43458 datasets were presented in the volcano plot using the ggplot2 package. FunRich software (14), a functional enrichment and interaction network analysis tool, was used to screen the overlapping genes among the TCGA-LUAD, GSE43458, and the two co-expression genes that came from the co-expression network. These genes were considered to be of prognostic value.

## Functional Enrichment Analysis

The clusterProlifer package (15) was used to perform Gene Ontology (GO) (16) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, which included biological processes, cellular components, molecular functions. Adjusted p-value <0.05 was set as the threshold value.

## PPI Network Construction

The Search Tool for the Retrieval of Interacting Genes(STRING, http://string-db.org) (17) was used to construct the interactive network of the overlapping DEGs and subsequently was visualized using the Cytoscape (18) software. The combined

score <0.4 was set as the cut-off criteria. In this study, we used the plugin CytoHubba (19) and MCODE to identify the hub genes and functional modules in the PPI network.

## Expression Level and Prognostic Value of Hub Genes

GEPIA (20) (http://gepia.cancerpku.cn/index.html) was used to study the gene expression of normal group and tumor groups based on different clinical stages. The Survival package was used to dig out the relationship between the overall survival (OS) and the hub genes based on TCGA data. We divided the hub genes into two groups based on their median expression value. The gene with the Log-rank test p < 0.05 was considered statistically significant. We also performed analyzed the correlation of gene expression and clinical parameters in LUAD. The immune infiltration analysis of hub genes was performed by TIMER (21) (https://cistrome.shinyapps.io/timer/).

## Cell Culture

The LUAD cell lines (H1299, H1975, and A549) were purchased from Chinese Academy of Sciences Cell Bank (Shanghai, China) and HBE cell was maintained in our laboratory. HBE served as control. All cells cultured in RPML-1640 medium (Gibco, USA) with 10% fetal bovine serum (Biowest S.A.S, France) and 1% penicillin-streptomycin (Gibco, USA) in a humidified incubator at 37°C containing 5% $CO_2$.

## RT-PCR

The TRIzol method was used to extract total RNA from cell lines, then transformed into cDNA by using SureScriptTM First-Strand cDNA Synthesis Kit (GeneCopoeia, Guangzhou, China). The expression level of SPP1 and VWF were measured by qRT-PCR using the commercial kit (BlazaTaqTM SYBR® Green qPCR Mix 2.0 (GeneCopoeia, Guangzhou, China) and a LightCycler 480 II System 373 (Roche, USA). β-actin set as the internal control. The primer sequences are list follows SPP1-F: AGCAGAATCTCCTAGCCCCA, SPP1-R: TGGTCATGGC TTTCGTTGGA; VWF-F: TTGTGGGAGATGTTTGCCTAC, VWF–R: CCTCTCTCATTGACCTTGCAG; ACTIN-F: CTT CGCGGGCGACGAT, ACTIN-R: CCATAGGAATCC TTCTGACC. The expression quantification was obtained with the $2^{-\Delta\Delta Ct}$ method.

## Western Blotting

Cells were lysed in IP lysis buffer containing protease inhibitors and then measured protein concentration by BCA assay. An equal amount of protein from cells loaded into the 8% SDS-PAGE, transferred to polyvinylidene difluoride membranes (PVDF), then incubated with primary antibody against VWF (1:500, Cell Signaling Technology, Shanghai, China) and SPP1 (1:2,000, Proteintech, Wuhan, China) at 4°C overnight, followed by incubation with peroxidase-conjugated secondary antibody (1:5,000, Abbkine, Wuhan, China) at room temperature for 1 h, finally visualized by enhanced chemiluminescence detection. The β-actin (1:5,000, Abbkine, Wuhan, China) was used as internal references in this study.

## Immunohistochemistry and Evaluation of Staining

Ten cases of fresh-frozen and paraffin-embedded LUAD tissues and adjacent non-cancerous tissues of LUAD patients were collected from the Department of Thoracic Surgery, Xiangya Hospital of Central South University, China. The research was approved by the local ethics committee (No. 202103046). The tissue was deparaffinized, rehydrated, and treated with an antigen retrieval in 0.01M citrate salt buffer (ZSGB-BIO, Beijing, China). After incubation of anti-VWF rabbit antibody (1:400, Abcam, Shanghai, China) overnight at 4°C, the tissue sections were incubated with horseradish peroxidase-conjugated anti-rabbit antibody (1:1,000, ZSGB-BIO, Beijing, China). Finally, the sections were stained with DAB+ substrate-chromogen solution (ZSGB-BIO, Beijing, China) and counterstained with hematoxylin. In negative controls, primary antibodies were omitted.

Two investigators assessed the immunostained result in a blinded fashion. The expression level of VWF was evaluated by intensity and percentage of positive cells according to a semi-quantitative method (22). The intensity was defined as 0 for negative staining, 1 for weak staining, 2 for moderate staining, and 3 for strong staining. Percentage of positive cells was quantified as 0 for ≤5% positive cells, 1 for 6–25%, 2 for 26–

50%, 3 for 51–75%, and 4 for ≥76%. The score was calculated by multiplying the score of staining intensity and percentage of positive cells.
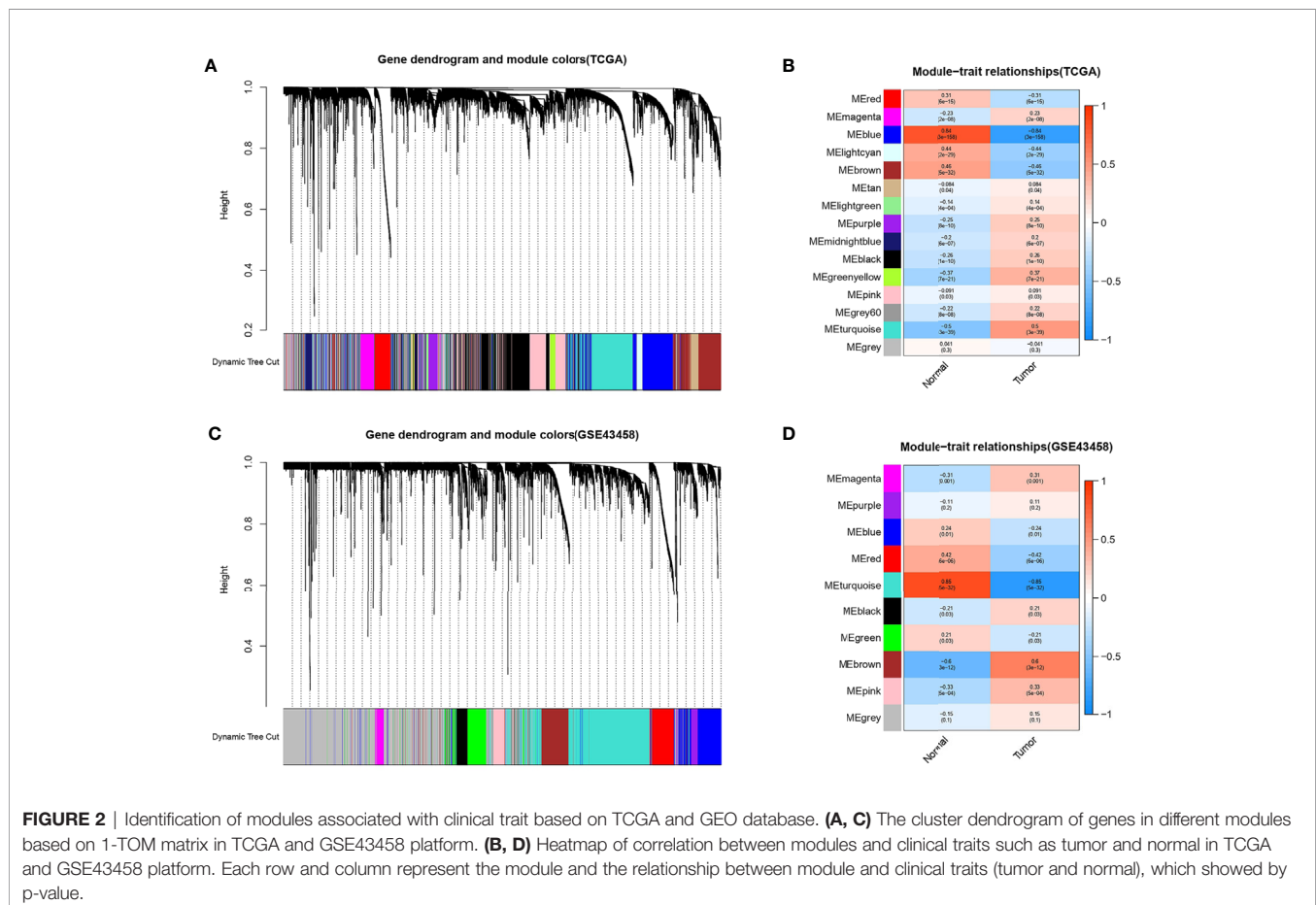
## Statistical Analysis

Statistical analyses were performed using GraphPad Prism 8 software. The result of RT-PCR was compared using the Students' t-test. Wilcoxon test was used to estimate the relationship between gene expression and LUAD patients' clinical parameters. The difference of VWF in LUAD tissue and normal lung tissues was analyzed by Mann–Whitney U-test. The p-value <0.05 was considered statistically significant.

# RESULTS

## Construction of Weighted Gene Co-expression Modules

The gene co-expression network was constructed from TCGA-LUAD and GSE43458 dataset using the WGCNA package. There were 14 and 9 modules identified in the TCGA-LUAD (**Figure 2A**) and the GSE43458 (**Figure 2C**). Then, the module-trait relationship was shown in the heatmap to estimate the relevance between each module and two clinical traits (tumor and normal).



**FIGURE 2** | Identification of modules associated with clinical trait based on TCGA and GEO database. **(A, C)** The cluster dendrogram of genes in different modules based on 1-TOM matrix in TCGA and GSE43458 platform. **(B, D)** Heatmap of correlation between modules and clinical traits such as tumor and normal in TCGA and GSE43458 platform. Each row and column represent the module and the relationship between module and clinical traits (tumor and normal), which showed by p-value.

As we can see from **Figures 2B, D**, the blue module (r = 0.84, p = 3e-158) in TCGA-LUAD and the turquoise module (r = 0.85, p = 5e-32) in GEO43458 had the strongest correlation with normal tissues. The genes in the TCGA blue module were mainly enriched in positive regulation of transcription from RNA polymerase II promoter (biological process), calcium ion binding (molecular function), and integral component of membrane (cell component). The genes in the GEO turquoise module were mainly enriched in cell adhesion (biological process), calcium ion binding (molecular function), and plasma membrane (cell component).

## Screening for Interest Differential Expression Genes

Setting the standard that adjusted p < 0.05 and the absolute of logFC larger than 1, we identified 3,582 (**Figure 3A**) and 799 (**Figure 3B**) DEGs in TCGA-LUAD and GSE32665. And 1,153 genes and 1,215 genes were identified in the TCGA-blue module and GEO- turquoise module. In addition, 366 overlapping genes were found by the Veen diagram (**Figure 3C**).
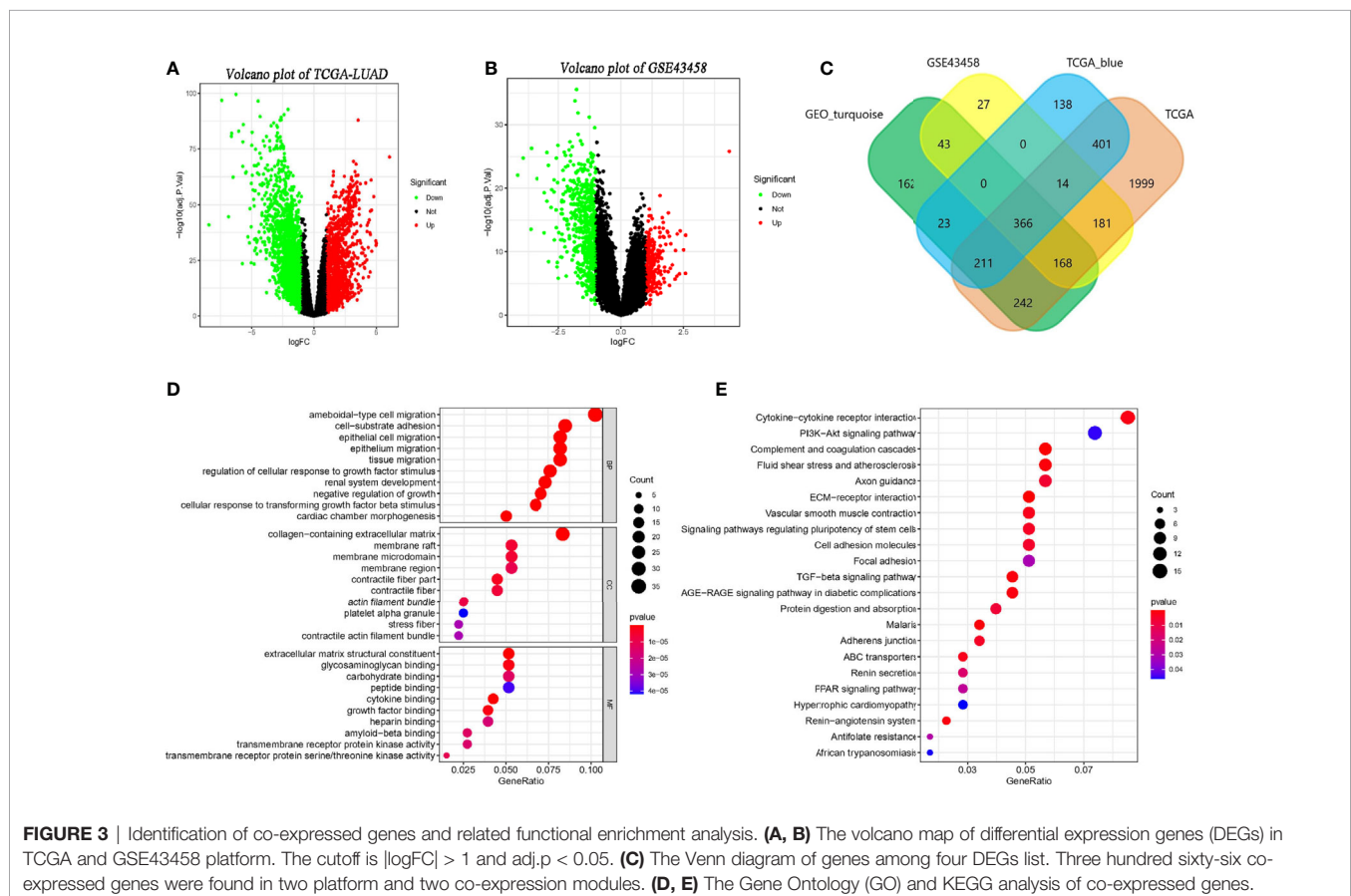
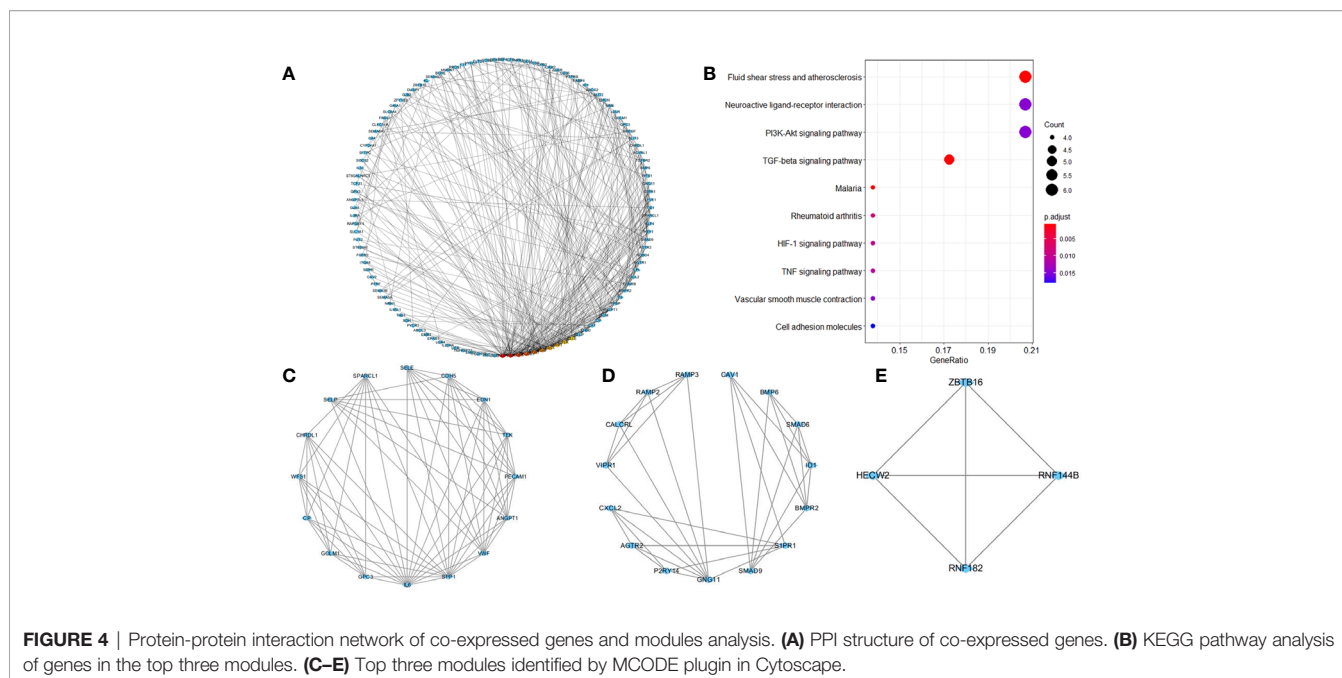## GO and KEGG Pathway Enrichment Analyses of Overlapping Genes

The GO and KEGG pathway analysis of 366 overlapping genes were processed by clusterProlifer package. In biological process

(BP) term, overlapping genes mainly in ameboidal-type cell migration, cell-substrate adhesion, and epithelial cell proliferation. As of cell component (CC) term, overlapping genes were extremely involved in the collagen-containing extracellular matrix, membrane raft, and membrane microdomain. In molecular function (MF) term, 366 genes were mainly associated with extracellular matrix structure constituent, glycosaminoglycan binding, and carbohydrate-binding (**Figure 3D**). KEGG analysis indicated those genes were mainly involved in cytokine-cytokine receptor interaction, PI3K-Akt signaling pathway, and complement and coagulation cascades (**Figure 3E**).

## Protein-Protein Interaction (PPI) Network Construction and Screening of Hub Genes

The PPI network of overlapping genes consisted of 365 nodes with 806 edges, and the average clustering coefficient was 0.367 (**Figure 4A**). The top 10 hub genes were screened by the CytoHubba according to their degree values, including IL6, VWF, CDH5, PECAM1, EDN1, BDNF, CAV1, SPP1, TEK, and SELE. A total of 17 modules were identified *via* the MCODE plugin and the top 3 modules are shown in (**Figures 4C–E**). KEGG pathway analysis showed that hub genes in the three modules were mainly associated with fluid shear stress and atherosclerosis, neuroactive ligand-receptor interaction, and PI3K-Akt signaling pathway (**Figure 4B**).



**FIGURE 3** | Identification of co-expressed genes and related functional enrichment analysis. **(A, B)** The volcano map of differential expression genes (DEGs) in TCGA and GSE43458 platform. The cutoff is |logFC| > 1 and adj.p < 0.05. **(C)** The Venn diagram of genes among four DEGs list. Three hundred sixty-six co-expressed genes were found in two platform and two co-expression modules. **(D, E)** The Gene Ontology (GO) and KEGG analysis of co-expressed genes.

**FIGURE 4** | Protein-protein interaction network of co-expressed genes and modules analysis. **(A)** PPI structure of co-expressed genes. **(B)** KEGG pathway analysis of genes in the top three modules. **(C–E)** Top three modules identified by MCODE plugin in Cytoscape.

## Expression Level and Prognostic Value of Hub Genes

From **Figure 5**, it can be seen that 9 of the 10 hub genes were lower expressed in lung adenocarcinoma compared with normal tissues, whereas expression of SPP1 was relatively higher expressed in the LUAD tissues. The OS analysis was performed using the survival package. As shown in **Figure 6**, the high expression of VWF was concerned with favorable prognosis, and the high expression of SPP1 was associated with worse prognosis ($p < 0.05$). Next, we performed RT-PCR, western blotting, and immunohistochemistry experiments to validate the mRNA and protein levels of SPP1 and VWF in LUAD cell lines and tissues. The result of qRT-PCR (**Figures 7A, B**) showed a higher level of SPP1 and a lower level of VWF in three LUAD cell lines compared with normal control (HBE). However, the western blotting analysis indicated compared with the low expression of VWF in three LUAD cell lines, SPP1 was highly expressed in only one cell line (A549), not in H1299 and H1975 cell lines (**Figure 7C**). As for tissue samples, Li group had validated the difference of SPP1 expression in LUAD patients using IHC (23), we just performed an analysis of VWF expression in IHC, the IHC analysis presented low expression level in LUAD tissue compared with normal lung tissue (**Figures 7D, E**).

We also extracted data from TCGA-LUAD and GSE43458 dataset to explore the relationship between hub gene expression and clinicopathological parameters in LUAD, the VWF and SPP1 expression were not significantly associated with gender, age, race, smoking habit, and stage (**Figure 8**). Furthermore, to dig out the function of hub genes in the immune microenvironment, we used the TIMER dataset to estimate the correlations of two hub genes with tumor purity and infiltrating immune cells (**Figure 9**). There was a weak relationship between

VWF, SPP1 expression, and infiltration of immune cells in LUAD (**Figures 9A, B**). VWF expression was associated with the infiltration level of B cell ($r = 0.219$, $p = 1.16e-06$), CD8+T cell ($r = 0.173$, p-1.30e-04), macrophage ($r = 0.275$, $p = 7.03e-10$), neutrophil ($r = 0.239$, $p = 1.01e-07$), and dendritic cell ($r = 0.235$, $p = 1.43e-07$) in LUAD. SPP1 expression level in LUAD was correlated with infiltrating CD4+T cell ($r = -0.076$, $p = 9.39e-02$), macrophage ($r = 0.292$, $p = 5.17e-11$), neutrophil ($r = 0.268$, $p = 2.09e-09$), and dendritic cell ($r = 0.291$, $p = 5.63e-11$).

## DISCUSSION

In recent decades, lung adenocarcinoma has become one of the most common malignancies in lung cancer. Most patients are diagnosed in the late stage because the early symptoms are not obvious. Through the treatment and diagnostic approaches have been improved, the mortality and incidence of lung adenocarcinoma remain high. Therefore, there is an urgent need to discover potential biomarkers for the prediction of the development and prognosis of LUAD.

In this study, WGCNA was used to construct the co-expression network of TCGA-LUAD and GSE43458 databases and identify relative gene modules. Three hundred sixty-six overlapping genes were found using integrated bioinformatics analysis. The overlapping genes were mainly involved in ameboidal-type cell migration (BP), collagen-containing extracellular matrix (CC), and extracellular matrix structural constituent (MF). Also, the plugin in Cytoscape, CytoHubba, and MCODE, were used to identify the hub genes and screen modules in the PPI network. The hub genes were IL6, VWF, CDH5, PECAM1, EDN1, BDNF, CAV1, SPP1, TEK, and SELE. Compared with normal tissues, only the SPP1 expression
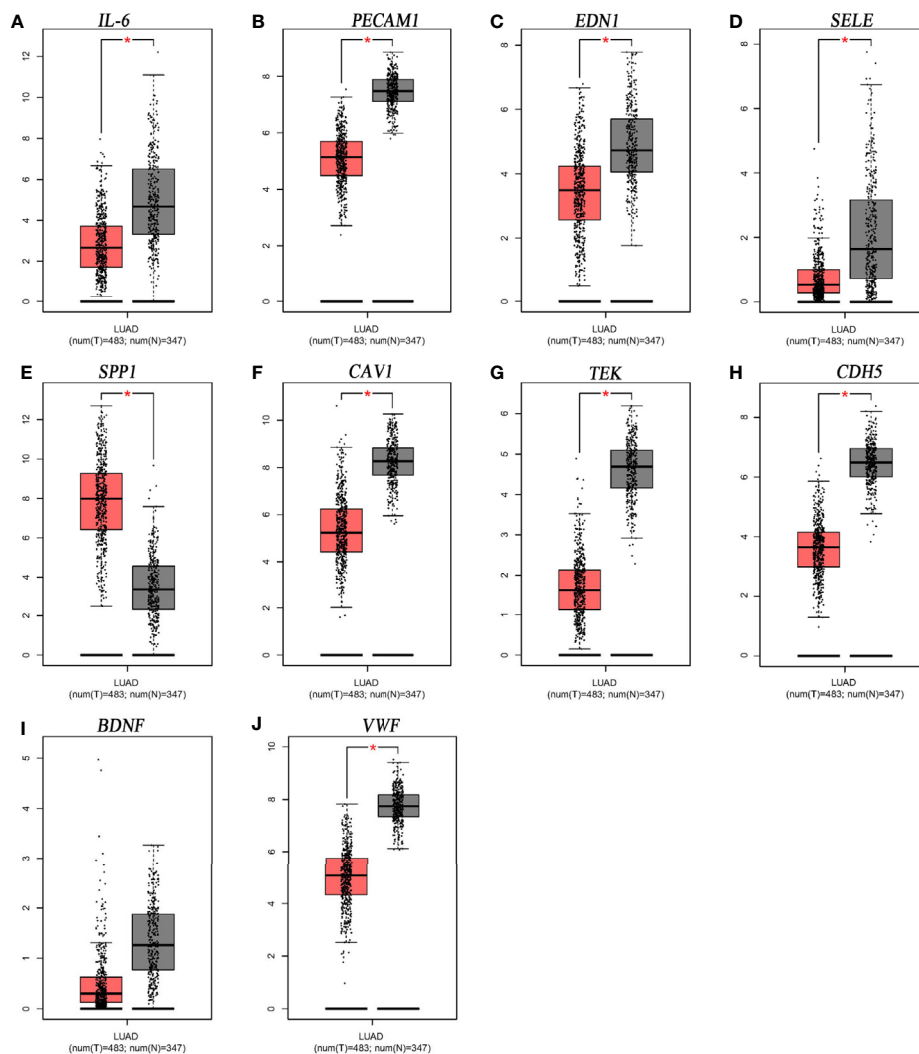
**FIGURE 5** | Expression level analysis of hub gens. **(A)** IL6; **(B)** PECAM1; **(C)** EDN1; **(D)** SELE; **(E)** SPP1; **(F)** CAV1; **(G)** TEK; **(H)** CDH5; **(I)** BDNF; **(J)** VWF. *p<0.05.

increased in LUAD tissues, while the rest hub genes decreased. The survival analysis indicated high expression of SPP1 was associated with poor overall survival in LUAD, while low expression of VWF was related to favorable overall survival. The protein expression level of SPP1 in H1299 and H1975 cell lines was inconsistent with mRNA level, which may be disturbed by post-translation, such as phosphorylation and ubiquitin.

Our workflow of the study was similar to the work (24). The other work screened the DEGs using differential gene expression analysis and identified hub genes in the PPI network, then performed related analysis to explore the relationship of hub genes and LUAD. The hub genes expression validated in RT-PCR and western blotting. CDH5, PECAM1, and VWF may play an important role in LUAD. We screened the co-expression genes using differential gene expression and WGCNA, identified hub genes from the PPI network, following the survival, and immune infiltration analysis, then expression validation using

RT-PCR, western blotting, and IHC. VWF and SPP1 can be used for further study. We all found the hub genes (IL6, VWF, CDH5, PECAM1, SPP1) from different PPI networks, and VWF and SPP1 expression were associated with LUAD patients' prognosis. But some differences need to be noticed. Firstly, the selection of data sources from the different platforms is more representative. Secondly, the idea of the study was different. WGCNA was powerful to find trait-related genes. Thirdly, immune infiltration analysis and many validation experiments have further enhanced the importance of hub genes. The same hub genes identified from different methods may play a role in the development of LUAD. SPP1, secreted phosphoprotein 1, also named osteopontin (OPN), is a secreted calcium-binding phosphorylation protein. SPP1 plays an important role in the regulation of biological processes such as cell proliferation, angiogenesis, inflammation, and apoptosis (25, 26). More evidence shows that SPP1 was overexpressed in various types
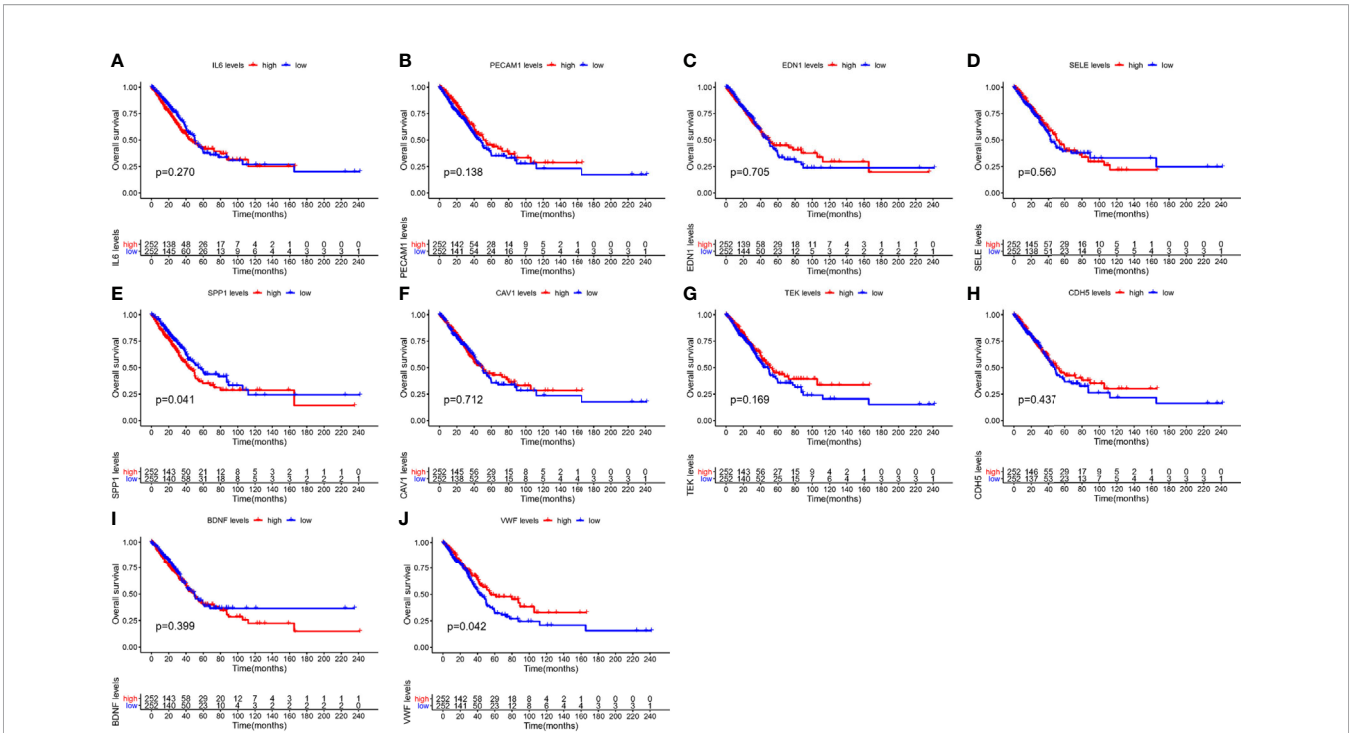
**FIGURE 6** | Survival analysis of hub gens. **(A)** IL6; **(B)** PECAM1; **(C)** EDN1; **(D)** SELE; **(E)** SPP1; **(F)** CAV1; **(G)** TEK; **(H)** CDH5; **(I)** BDNF; **(J)** VWF.

of cancers and contributes to the occurrence and development of tumors, including colorectal cancer, cervical cancer, gastric cancer, and so on (27–30). The result of our finding that SPP1 may be a potential biomarker for lung adenocarcinoma was consistent with several studies (31, 32). More reports had

illustrated the relationship between the expression level of SPP1 and lung adenocarcinoma. For example, by inhibiting autophagy and apoptosis, SPP1 promotes the development of small cell lung cancer (33). SPP1 enhanced the second-generation epidermal growth factor receptor (EGFR) tyrosine
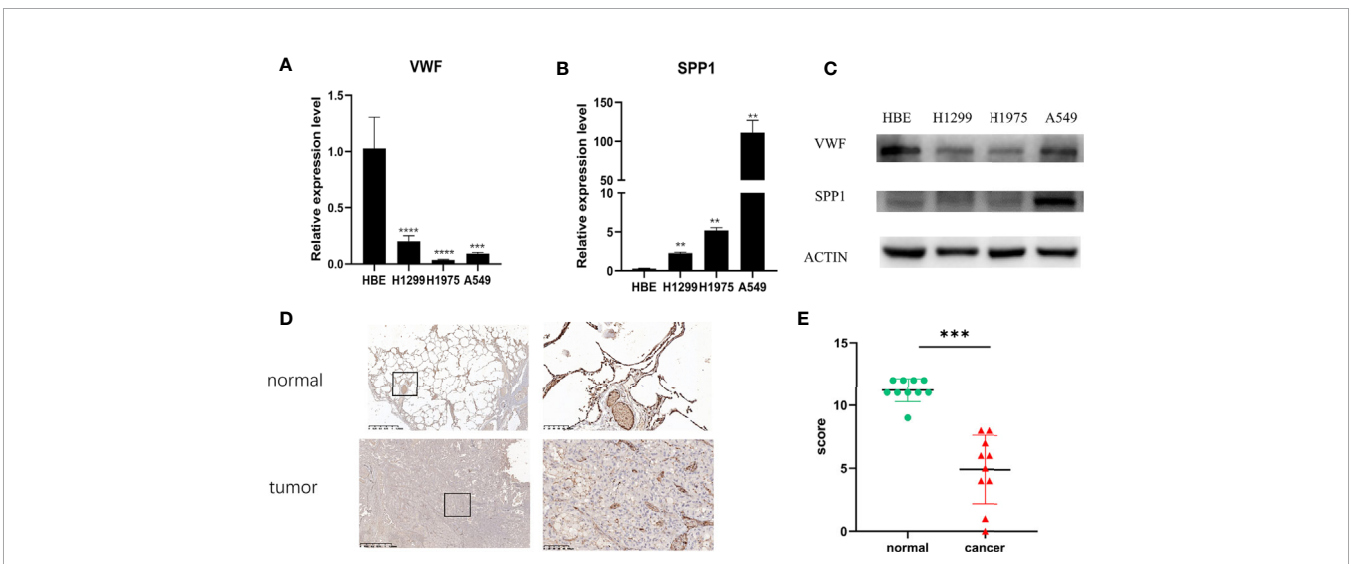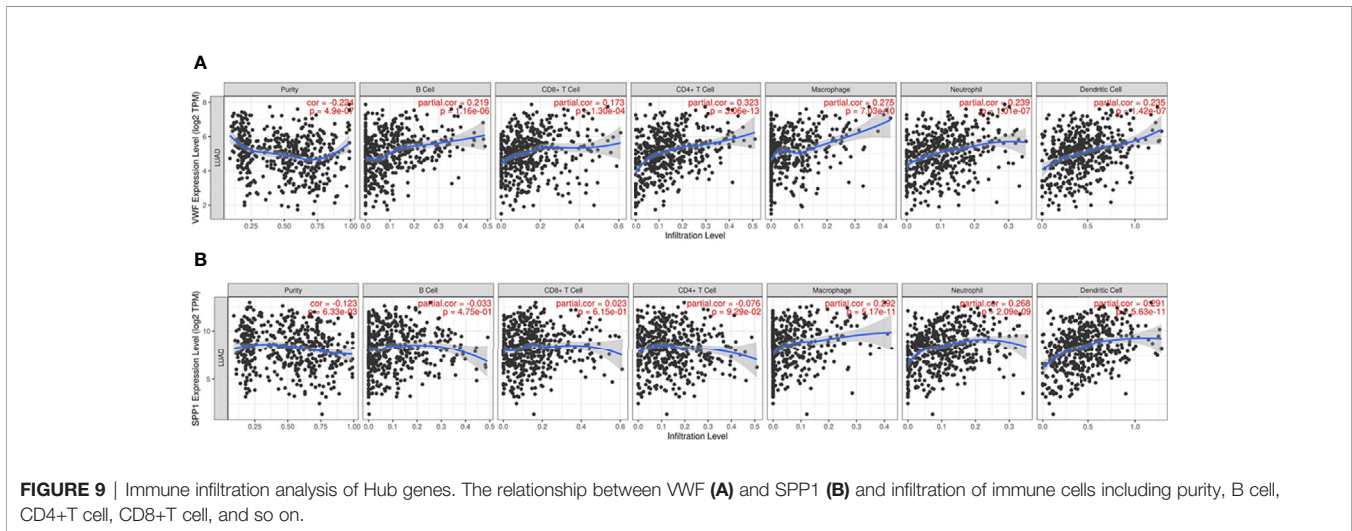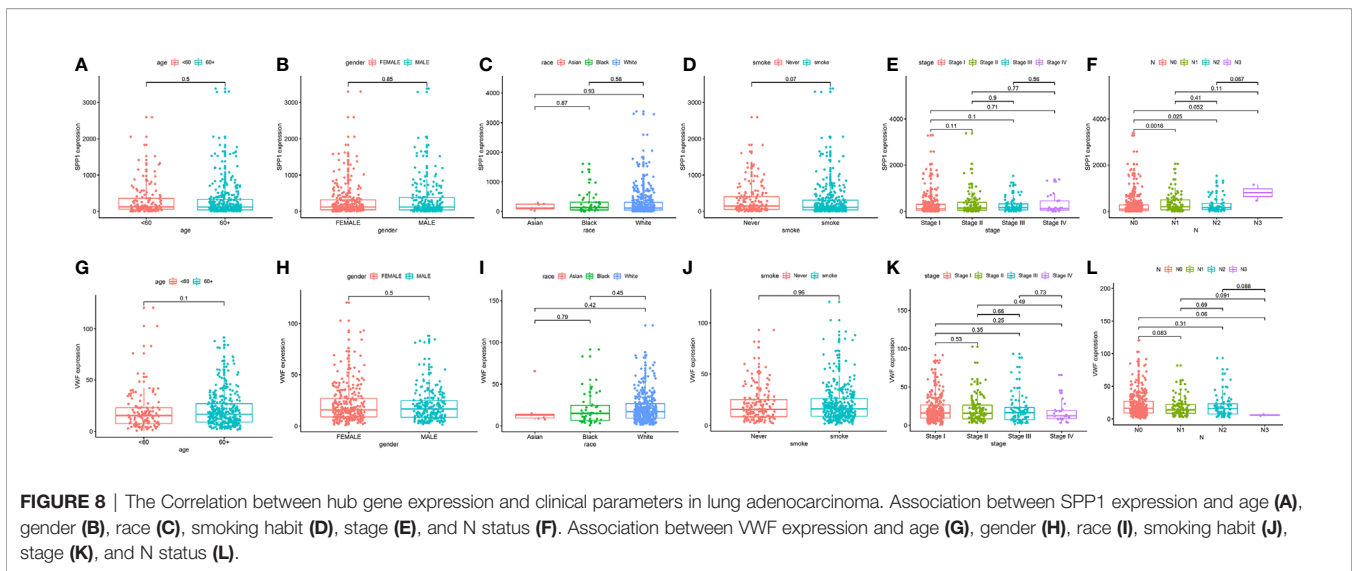


**FIGURE 7** | Validation of mRNA and protein expression level. The mRNA expression level of VWF **(A)** and SPP1 **(B)** in HBE, H1299, H1975, and A549 cell lines. **p<0.01,***p<0.001,****p<0.0001. And protein expression of VWF and SPP1 validated using western blotting **(C)**. Each experiment was carried out three times. A representative immunohistochemistry image of VWF in LUAD and adjacent non-tumor tissues. Scale bar = 100 µm **(D)**. The scatter plot of immunohistochemistry score of LUAD and matched normal tissues. **(E)**. ***p< 0.001.

**FIGURE 8** | The Correlation between hub gene expression and clinical parameters in lung adenocarcinoma. Association between SPP1 expression and age **(A)**, gender **(B)**, race **(C)**, smoking habit **(D)**, stage **(E)**, and N status **(F)**. Association between VWF expression and age **(G)**, gender **(H)**, race **(I)**, smoking habit **(J)**, stage **(K)**, and N status **(L)**.



**FIGURE 9** | Immune infiltration analysis of Hub genes. The relationship between VWF **(A)** and SPP1 **(B)** and infiltration of immune cells including purity, B cell, CD4+T cell, CD8+T cell, and so on.

kinase inhibitor (TKI) resistance in lung cancer, which offered a new scheme for targeted therapy that inhibiting SPP1 might be a therapeutic target to overcome afatinib resistance (34). Meanwhile, SPP1 mediating macrophage polarization leads to upregulation of PD-L1 and facilitates immune escape in lung adenocarcinoma, which also suggests a potential therapeutic target for lung cancer (35). These studies further presented the importance of SPP1 in the mechanism of metastasis and invasion of the tumor.

VWF, von Willebrand factor, a multimeric protein produced by endothelial cells, platelets, and megakaryocytes, dedicates to primary hemostasis, mediating platelets adhesion and acting as a carrier protein for coagulation factor (F) VIII (36, 37). The VWF is highly expressed in cancers such as kidney renal clear cell carcinoma, glioblastoma multiforme, and liver hepatocellular carcinoma, while lowly expressed in tumors including colon adenocarcinoma, breast invasive carcinoma, and uterine corpus endometrial carcinoma. Due to the significance of hemostasis,

VWF plays a role in cancers by inhibiting angiogenesis and apoptosis (38–40). The high expression of VWF relates to endothelial damage and is used for indicating many pathological situations, including atherosclerosis, cardiovascular diseases, and cancers (41–44). Besides, scientists pay attention to explore the relationship between the expression level of VWF and lung cancer. Liu group found that VWF expression in endothelial cells of LUAD is elevated by the cancer cell-derived secretome through GATA3-mediated transcription (45); According to a study, which digs out the association between VWF and its cleaving protease (ADAMTS-13) in lung cancer metastasis, the elevation of VWF/ADAMTS-13 ratio may serve as an independent predictive factor for mortality in patients with advanced NSCLC (46). Another useful study considered the interference of different blood groups while studying the expression levels of VWF and ADAMTS-13 in lung cancer patients, showed that increased VWF and decreased ADAMTS-13 promote the invasion of lung cancer and non-O

blood group is a risk factor for increased VWF and FVIII in plasma (47). The special molecular mechanism of how VWF affects lung adenocarcinoma is still underway. Thus, further studies are needed to confirm this contention.

Although we have preliminary screened the potential prognostic biomarkers of LUAD, there still exist some limitations. On one hand, the number of samples is insufficient to avoid bias, on the other hand, more in-depth studies are needed in the future. Our analysis can provide valuable information for researchers to find possible genes and pathways concerned with LUAD for better understanding the molecular mechanism of LUAD carcinogenesis.

In summary, we identified 10 hub genes by integrating the GEO and TCGA dataset with WGCNA, and further showed that VWF down-regulation might be a prognostic biomarker in LUAD. These findings reported here could have clinical value in prognosis of LUAD, and provide valuable information for further studying the roles of these differential hub genes in LUAD.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/ and https://www.ncbi.nlm.nih.gov/gds.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Xiangya Hospital of Central South University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LZ, SZ, and WB: design and methodology. XL and YH: software and data procession. JJ and YH: validation. YH, RL, MY, and WB: writing-original draft preparation. XL and PZ: writing-review, editing, and supervision. All authors contributed to the article and approved the submitted version.

## FUNDING

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.639600/full#supplementary-material

## REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA Cancer J Clin* (2015) 65(2):87–108. doi: 10.3322/caac.21262

2. Martin P, Leighl NB. Review of the Use of Pretest Probability for Molecular Testing in non-Small Cell Lung Cancer and Overview of New Mutations That may Affect Clinical Practice. *Ther Adv Med Oncol* (2017) 9(6):405–14. doi: 10.1177/1758834017704329

3. Liu J, Li Y, Zhu X, Li Q, Liang X, Xie J, et al. Increased CX3CL1 mRNA Expression Level is a Positive Prognostic Factor in Patients With Lung Adenocarcinoma. *Oncol Lett* (2019) 17(6):4877–90. doi: 10.3892/ol.2019.10211

4. Bai Y, Xiong L, Zhu M, Yang Z, Zhao J, Tang H. Co-Expression Network Analysis Identified KIF2C in Association With Progression and Prognosis in Lung Adenocarcinoma. *Cancer Biomark* (2019) 24(3):371–82. doi: 10.3233/CBM-181512

5. Long HP, Liu JQ, Yu YY, Qiao Q, Li G. PKMYT1 as a Potential Target to Improve the Radiosensitivity of Lung Adenocarcinoma. *Front Genet* (2020) 11:376. doi: 10.3389/fgene.2020.00376

6. 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

7. Niemira M, Collin F, Szalkowska A, Bielska A, Chwialkowska K, Reszec J, et al. Molecular Signature of Subtypes of Non-Small-Cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-Expression Network Analysis (Wgcna). *Cancers (Basel)* (2019) 12(1):37. doi: 10.3390/cancers12010037

8. Yang Q, Wang R, Wei B, Peng C, Wang L, Hu G, et al. Candidate Biomarkers and Molecular Mechanism Investigation for Glioblastoma Multiforme Utilizing Wgcna. *BioMed Res Int* (2018) 2018:4246703. doi: 10.1155/2018/4246703

9. Li J, Zhou D, Qiu W, Shi Y, Yang JJ, Chen S, et al. Application of Weighted Gene Co-expression Network Analysis for Data From Paired Design. *Sci Rep* (2018) 8(1):622. doi: 10.1038/s41598-017-18705-z

10. Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H. Gene Co-Expression Network Analysis Reveals Common System-Level Properties of Prognostic Genes Across Cancer Types. *Nat Commun* (2014) 5:3231. doi: 10.1038/ncomms4231

11. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* (2010) 26(1):139–40. doi: 10.1093/bioinformatics/btp616

12. 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

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. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. Funrich: An Open Access Standalone Functional Enrichment and Interaction Network Analysis Tool. *Proteomics* (2015) 15(15):2597–601. doi: 10.1002/pmic.201400515

15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

16. Gene Ontology C. The Gene Ontology (GO) Project in 2006. *Nucleic Acids Res* (2006) 34(Database issue):D322–326. doi: 10.1093/nar/gkj021

17. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. String v11: Protein-Protein Association Networks With Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res* (2019) 47(D1):D607–13. doi: 10.1093/nar/gky1131

18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303

19. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: Identifying Hub Objects and Sub-Networks From Complex Interactome. *BMC Syst Biol* (2014) 8(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11

20. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A Web Server for Cancer and Normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res* (2017) 45(W1):W98–W102. doi: 10.1093/nar/gkx247

21. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. Timer: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* (2017) 77(21):e108–10. doi: 10.1158/0008-5472.CAN-17-0307

22. Detre S, Saclani Jotti G, Dowsett M. A "Quickscore" Method for Immunohistochemical Semiquantitation: Validation for Oestrogen Receptor in Breast Carcinomas. *J Clin Pathol* (1995) 48(9):876–8. doi: 10.1136/jcp.48.9.876

23. Li S, Yang R, Sun X, Miao S, Lu T, Wang Y, et al. Identification of SPP1 as a Promising Biomarker to Predict Clinical Outcome of Lung Adenocarcinoma Individuals. *Gene* (2018) 679:398–404. doi: 10.1016/j.gene.2018.09.030

24. Fei H, Chen S, Xu C. Interactive Verification Analysis of Multiple Sequencing Data for Identifying Potential Biomarker of Lung Adenocarcinoma. *BioMed Res Int* (2020) 2020:8931419. doi: 10.1155/2020/8931419

25. Hubbard NE, Chen QJ, Sickafoose LK, Wood MB, Gregg JP, Abrahamsson NM, et al. Transgenic Mammary Epithelial Osteopontin (spp1) Expression Induces Proliferation and Alveologenesis. *Genes Cancer* (2013) 4(5-6):201–12. doi: 10.1177/1947601913496813

26. Lamort AS, Giopanou I, Psallidas I, Stathopoulos GT. Osteopontin as a Link Between Inflammation and Cancer: The Thorax in the Spotlight. *Cells* (2019) 8(8):815. doi: 10.3390/cells8080815

27. Choe EK, Yi JW, Chai YJ, Park KJ. Upregulation of the Adipokine Genes ADIPOR1 and SPP1 is Related to Poor Survival Outcomes in Colorectal Cancer. *J Surg Oncol* (2018) 117(8):1833–40. doi: 10.1002/jso.25078

28. Chen X, Xiong D, Ye L, Yang H, Mei S, Wu J, et al. SPP1 Inhibition Improves the Cisplatin Chemo-Sensitivity of Cervical Cancer Cell Lines. *Cancer Chemother Pharmacol* (2019) 83(4):603–13. doi: 10.1007/s00280-018-3759-5

29. Zhuo C, Li X, Zhuang H, Tian S, Cui H, Jiang R, et al. Elevated THBS2, COL1A2, and SPP1 Expression Levels as Predictors of Gastric Cancer Prognosis. *Cell Physiol Biochem* (2016) 40(6):1316–24. doi: 10.1159/000453184

30. Arendt BM, Teterina A, Pettinelli P, Comelli EM, Ma DWL, Fung SK, et al. Cancer-Related Gene Expression is Associated With Disease Severity and Modifiable Lifestyle Factors in non-Alcoholic Fatty Liver Disease. *Nutrition* (2019) 62:100–7. doi: 10.1016/j.nut.2018.12.001

31. Zhang W, Fan J, Chen Q, Lei C, Qiao B, Liu Q. SPP1 and AGER as Potential Prognostic Biomarkers for Lung Adenocarcinoma. *Oncol Lett* (2018) 15(5):7028–36. doi: 10.3892/ol.2018.8235

32. Guo Z, Huang J, Wang Y, Liu XP, Li W, Yao J, et al. Analysis of Expression and Its Clinical Significance of the Secreted Phosphoprotein 1 in Lung Adenocarcinoma. *Front Genet* (2020) 11:547. doi: 10.3389/fgene.2020.00547

33. Liu H, Wei S, Zhang L, Yuan C, Duan Y, Wang Q. Secreted Phosphoprotein 1 Promotes the Development of Small Cell Lung Cancer Cells by Inhibiting Autophagy and Apoptosis. *Pathol Oncol Res* (2019) 25(4):1487–95. doi: 10.1007/s12253-018-0504-7

34. Wang X, Zhang F, Yang X, Xue M, Li X, Gao Y, et al. Secreted Phosphoprotein 1 (Spp1) Contributes to Second-Generation Egfr Tyrosine Kinase Inhibitor Resistance in Non-Small Cell Lung Cancer. *Oncol Res* (2019) 27(8):871–7. doi: 10.3727/096504018X15426271404407

35. Zhang Y, Du W, Chen Z, Xiang C. Upregulation of PD-L1 by SPP1 Mediates Macrophage Polarization and Facilitates Immune Escape in Lung Adenocarcinoma. *Exp Cell Res* (2017) 359(2):449–57. doi: 10.1016/j.yexcr.2017.08.028

36. Terraube V, O'Donnell JS, Jenkins PV. Factor VIII and Von Willebrand Factor Interaction: Biological, Clinical and Therapeutic Importance. *Haemophilia* (2010) 16(1):3–13. doi: 10.1111/j.1365-2516.2009.02005.x

37. Sporn LA, Marder VJ, Wagner DD. Inducible Secretion of Large, Biologically Potent Von Willebrand Factor Multimers. *Cell* (1986) 46(2):185–90. doi: 10.1016/0092-8674(86)90735-x

38. Franchini M, Frattini F, Crestani S, Bonfanti C, Lippi G. Von Willebrand Factor and Cancer: A Renewed Interest. *Thromb Res* (2013) 131(4):290–2. doi: 10.1016/j.thromres.2013.01.015

39. Borsig L. VWF Fibers Induce Thrombosis During Cancer. *Blood* (2015) 125(20):3042–3. doi: 10.1182/blood-2015-03-632141

40. Pepin M, Kleinjan A, Hajage D, Buller HR, Di Nisio M, Kamphuisen PW, et al. Adamts-13 and Von Willebrand Factor Predict Venous Thromboembolism in Patients With Cancer. *J Thromb Haemost* (2016) 14(2):306–15. doi: 10.1111/jth.13205

41. Wagner DD, Frenette PS. The Vessel Wall and its Interactions. *Blood* (2008) 111(11):5271–81. doi: 10.1182/blood-2008-01-078204

42. Lehrer S, Green S, Dembitzer FR, Rheinstein PH, Rosenzweig KE. Increased RNA Expression of Von Willebrand Factor Gene is Associated With Infiltrating Lobular Breast Cancer and Normal Pam50 Subtype. *Cancer Genomics Proteomics* (2019) 16(3):147–53. doi: 10.21873/cgp.20120

43. Takaya H, Namisaki T, Kitade M, Kaji K, Nakanishi K, Tsuji Y, et al. VWF/ADAMTS13 Ratio as a Potential Biomarker for Early Detection of Hepatocellular Carcinoma. *BMC Gastroenterol* (2019) 19(1):167. doi: 10.1186/s12876-019-1082-1

44. Lehrer S, Rheinstein PH, Rosenzweig KE. Increased Expression of Von Willebrand Factor Gene is Associated With Poorer Survival in Primary Lower Grade Glioma. *Glioma* (2018) 1(4):132–5. doi: 10.4103/glioma.glioma_17_18

45. Xu Y, Pan S, Liu J, Dong F, Cheng Z, Zhang J, et al. GATA3-Induced vWF Upregulation in the Lung Adenocarcinoma Vasculature. *Oncotarget* (2017) 8(66):110517–29. doi: 10.18632/oncotarget.22806

46. Guo R, Yang J, Liu X, Wu J, Chen Y. Increased Von Willebrand Factor Over Decreased ADAMTS-13 Activity is Associated With Poor Prognosis in Patients With Advanced non-Small-Cell Lung Cancer. *J Clin Lab Anal* (2018) 32(1):e22219. doi: 10.1002/jcla.22219

47. Liu X, Chen X, Yang J, Guo R. Association of ABO Blood Groups With Von Willebrand Factor, Factor VIII and ADAMTS-13 in Patients With Lung Cancer. *Oncol Lett* (2017) 14(3):3787–94. doi: 10.3892/ol.2017.6619