



Research Article

Identification of Potential Prognostic Biomarkers Associated with Monocyte Infiltration in Lung Squamous Cell Carcinoma

Hailin Liu ^{1,2,3,4} Bo Yan,⁵ Yulong Chen,^{1,2,3,4} Juan Pang,^{2,3,4,6} Yue Li,^{1,2,3,4} Zhenfa Zhang,^{1,2,3,4} Chenguang Li,^{1,2,3,4} and Tingting Qin ^{2,3,4,7}

¹Department of Lung Cancer Tianjin Medical University Cancer Institute and Hospital, China

²Tianjin Medical University Cancer Institute and Hospital National Clinical Research Center for Cancer, China

³Key Laboratory of Cancer Prevention and Therapy, China

⁴Tianjin's Clinical Research Center for Cancer, China

⁵Department of Radiotherapy, Tianjin Medical University Cancer Institute and Hospital, China

⁶Department of Pathology Tianjin Medical University Cancer Institute and Hospital, China

⁷Department of Thoracic Oncology Tianjin Lung Cancer Center, Tianjin Cancer Institute and Hospital, China

Correspondence should be addressed to Tingting Qin; kualesaga@126.com

Received 10 March 2022; Revised 26 May 2022; Accepted 22 June 2022; Published 11 August 2022

Academic Editor: Surendiran B

Copyright © 2022 Hailin Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The five-year survival rate of lung squamous cell carcinoma is significantly lower than that of other cancer types. It is therefore urgent to discover novel prognosis biomarkers and therapeutic targets and understand their correction with infiltrating immune cells to improve the prognosis of patients with lung squamous cell carcinoma. In this study, we employed robust rank aggregation algorithms to overcome the shortcomings of small sizes and potential bias in each Gene Expression Omnibus dataset of lung squamous cell carcinoma and identified 513 robust differentially expressed genes including 220 upregulated and 293 downregulated genes from six microarray datasets. Functional enrichment analysis showed that these robust differentially expressed genes were obviously involved in the extracellular matrix and structure organization, epidermis development, cell adhesion molecule binding, p53 signaling pathway, and interleukin-17 signaling pathway to affect the progress of lung squamous cell carcinoma. We further identified six hub genes from 513 robust differentially expressed genes by protein-protein interaction network and 10 topological analyses. Moreover, the results of immune cell infiltration analysis from six integrated Gene Expression Omnibus datasets and our sequencing transcriptome data demonstrated that the abundance of monocytes was significantly lower in lung squamous cell carcinoma compared to controls. Immune correlation analysis and survival analysis of hub genes suggested that three hub genes, collagen alpha-1(VII) chain, mesothelin, and chordin-like protein 1, significantly correlated with tumor-infiltrating monocytes as well as may be potential prognostic biomarkers and therapy targets in lung squamous cell carcinoma. The investigation of the correlation of hub gene markers and infiltrating monocytes can also improve to well understand the molecular mechanisms of lung squamous cell carcinoma development.

1. Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide [1] that, even in the USA, still causes approximately 350 deaths each day [2]. Lung squamous cell carcinoma (LUSC), a type of non-small-cell lung cancer (NSCLC), is the second most common histological type of

lung cancer [3]. LUSC originates from the transformation of the squamous cells lining the central part of the lung or in the main airway, and it is more strongly associated with smoking [4]. The five-year survival rate of NSCLC is 15% [5], which is obviously unsatisfactory compared to other leading types of cancer in this world [6–8]. So, it is important to improve the survival rate and life quality of patients

TABLE 1: Characteristics and DEGs of GEO LUSC datasets.

Datasets	Platform	Characteristics of samples			Up	DEGs [#]	
		Noncancer	Cancer	Total		Down	Total
GSE1987	GPL91	9	17	26	215	255	470
GSE2088	GPL962	30	48	78	142	242	384
GSE8569	GPL5645	6	36	42	125	197	322
GSE21933	GPL6254	11	11	22	1683	1818	3501
GSE33479	GPL6480	13	14	27	250	180	430
GSE33532	GPL570	4	4	8	1330	1645	2975

[#]The cutoff criteria were $|\log_2 \text{FC}| > 1$ and adjusted $p < 0.05$ for DEGs analysis.

and discover new prognosis biomarkers and therapeutic targets for LUSC, and it will be beneficial to understand the development of LUSC.

Tumor microenvironment (TME) was infiltrated with multiple subtypes of immune cells, which could influence the survival rate and clinical characteristics [9, 10]. Hence, tumor-infiltrating immune cells (TICs) can be assumed to be important indicators to estimate the prognosis and therapeutic response [11, 12]. Cumulative evidences have indicated the correlation of tumor-infiltrating immune cells between hub genes involved in tumorigenesis [13, 14]. For example, hub genes of COL1A1, COL4A1, COL12A1, and PDGFRB were identified as potential prognostic biomarkers associated with macrophage M2 infiltration in gastric cancer, which play crucial roles in the proliferation and invasion of carcinoma cells [15]. Monocytes are mononuclear phagocytes [16], which play an important role in regulating tumorigenesis and metastasis [17]. The abundance of monocyte indicates clinicopathological characteristics and prognostic quality in NSCLC [18]. Some hub genes were significantly correlated with the scores of stromal, immune, and tumor purity [19]; however, it is unknown which hub genes were identified as potential prognostic biomarkers associated with monocyte infiltration in LUSC.

Many transcriptome microarray datasets on LUSC have accumulated in public databases, such as Gene Expression Omnibus (GEO) [20], which can be used to discover disease-related genes. Some scientists previously attempted to use some datasets to identify hub genes of LUSC with the knowledge from analyses [21, 22]. However, limited sample sizes could give rise to biased outcomes. In order to efficiently take advantage of mining GEO datasets, the robust rank aggregation (RRA) method was used to analyze more samples of multiple datasets [23], which could avoid the sample heterogeneity and the biases derived from different technology platforms [24, 25].

Herein, we integrated six GEO datasets with low errors and noise and identified 513 robust differentially expressed genes (DEGs) employing robust rank aggregation (RRA) algorithms. Integrative bioinformatics analyses were executed on robust DEGs to identify the molecular diagnosis markers and pathogenesis mechanisms for LUSC. Finally, COL7A1, MSLN, and CHRDL1 were considered potential prognostic biomarkers associated with LUSC-infiltrated monocytes, indicating the important roles of these hub genes in the pathogenesis of LUSC.

2. Methods

2.1. Tumor Samples and RNA Sequencing. Five paired LUSC and normal tissue samples were collected from patients by surgical operation at Tianjin Medical University Cancer Institute & Hospital. These tissue samples were snap-frozen with liquid nitrogen and stored at -80°C . The studies have been approved by the Ethics Committee of Tianjin Medical University Cancer Institute & Hospital and have signed informed consent from participants. The samples were sequenced by Novogene company (Tianjin, China). The matrices of count, FPKM, and TPM were used for further analysis.

2.2. Data Availability and Processing. Six microarray datasets, GSE1987, GSE2088, GSE8569, GSE21933, GSE33479, and GSE33532, were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo>). We collected a total of 203 samples consisting of 73 normal lung tissues and 130 LUSC tissues from the above 6 datasets. The detailed information of each dataset is described in Table 1. The probe names in each matrix file from six GEO datasets were replaced with gene symbols of the corresponding platform annotation document by Perl. Data normalization and DEGs in each dataset were carried out using the “limma” package in R software. Volcano plots of DEGs were displayed using the R package “ggplot2.” The cutoff criteria were $|\log_2 \text{FC}| > 1$ and adjusted p value < 0.05 .

2.3. Integrate Analysis through Robust Rank Aggregation (RRA) Method. We employed the RRA method [23] to integrate the above matrices, which could minimize the bias and errors among 6 microarray datasets. First, the upregulated and downregulated genes were categorized using the “limma” package in each dataset. Then, the categorized genes were conducted by the “RobustRankAggreg” package to filtrate robust DEGs. A heatmap of the top 20 robust DEGs were drawn using the “pheatmap” package. The cutoff criteria were $|\log_2 \text{FC}| > 1$ and $\text{FDR} < 0.05$.

2.4. GO Function and KEGG Pathway Enrichment Analyses. GO, KEGG, and DO analyses for robust DEGs were implemented using R package “<https://org.hs-eg.db>,” “enrichplot,” “DOSE,” “clusterProfiler,” and “ggplot2” to enrich biological process (BP), cellular component (CC), molecular function (MF), and signal pathways. The statistical significance was considered with a q value < 0.05 . Further, the

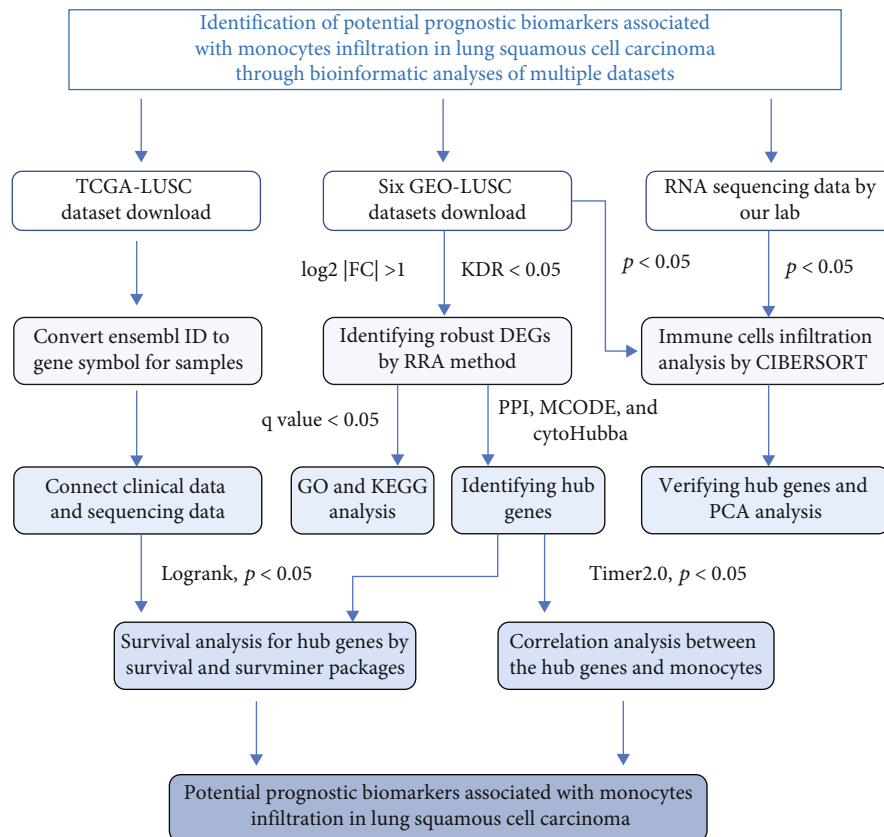


FIGURE 1: Schematic illustration of the bioinformatics analysis of multiple datasets.

important terms of GO and KEGG were investigated for overlapping robust DEGs through the “GPlot” package.

2.5. Protein-Protein Interaction Network Construction and Key Module Analysis. The robust DEGs were imported into the STRING online database (<http://www.string-db.org>) to construct the PPI network with confidence > 0.9 as cutoff criteria. Then, the PPI network was visualized by Cytoscape V3.8.0 [26]. An MCODE plugin was used to screen the key modules from the whole PPI network.

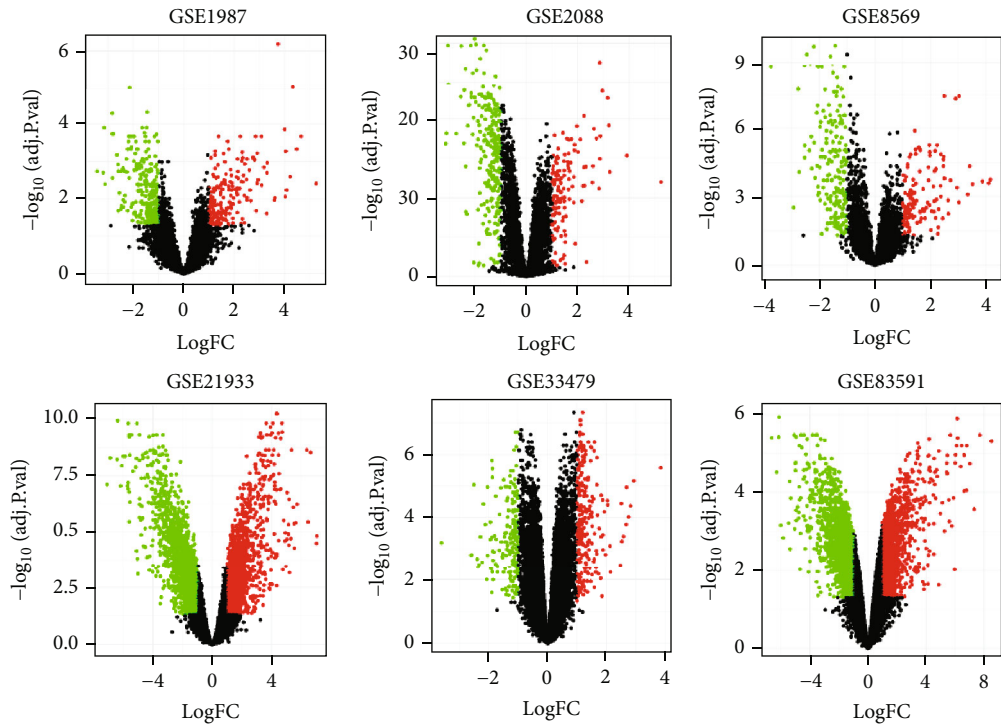
2.6. Hub Gene Identification. The cytoHubba plugin of Cytoscape including ten topological analysis algorithms, MCC, DMNC, MNC, Degree, EPC, BottleNeck, EcCentricity, Closeness, Radiality, and Betweenness, was employed to determine the hub genes in the whole PPI network [27]. The scores of each node in the whole PPI network were calculated to screen out hub genes from the top 100 nodes. The upset plot of hub genes was plotted using the “UpSetR” package.

2.7. Immune Cell Infiltration Prediction by CIBERSORT Analysis. The CIBERSORT algorithm was a deconvolution method based on the standardized gene expression profiles to predict the relative components of 22 subtypes of immune cells in tissue samples [28]. The standardized gene expression matrices from the six datasets were converted to the

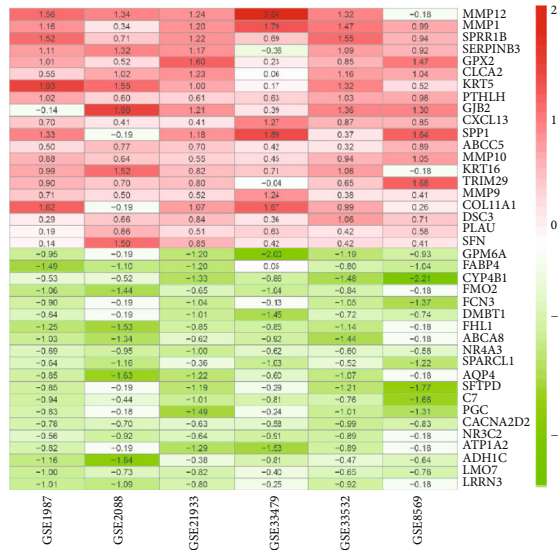
22 types of immune cell matrices referred to the leukocyte signature matrix (LM22). The cutoff criteria of p value < 0.05 for each sample indicates that the predicted proportion of each infiltrating immune cell subtype is significantly accurate and suitable for further analysis. The heatmap and violin plots of 22 subtypes of immune cells were drawn using “pheatmap” and “vioplot” packages. Principal component analysis (PCA) based on CIBERSORT-calculated results was used to determine the difference between normal and LUSC samples.

2.8. Correlation Analysis between the Hub Genes and Immune Cell Infiltration. First, the mRNA expression of hub genes was verified in TCGA and GTEX databases by GEPIA2 (<http://gepia2.cancer-pku.cn/>). Then, the mRNA expression of hub genes was further determined in RNA sequencing data from collected tissue samples. Finally, tumor-infiltrated monocytes were estimated by immune deconvolution methods on Timer 2.0 to investigate the relationship with hub genes’ expression level [29].

2.9. Survival Analysis. To investigate the overall survival of LUSC patients, TCGA-LUSC gene expression profiles and corresponding clinical data were downloaded from TCGA database (<https://portal.gdc.cancer.gov/>). Then, the ensembl ID was converted to a gene symbol by Perl. Next, we connected clinical data with the hub gene expression matrix.

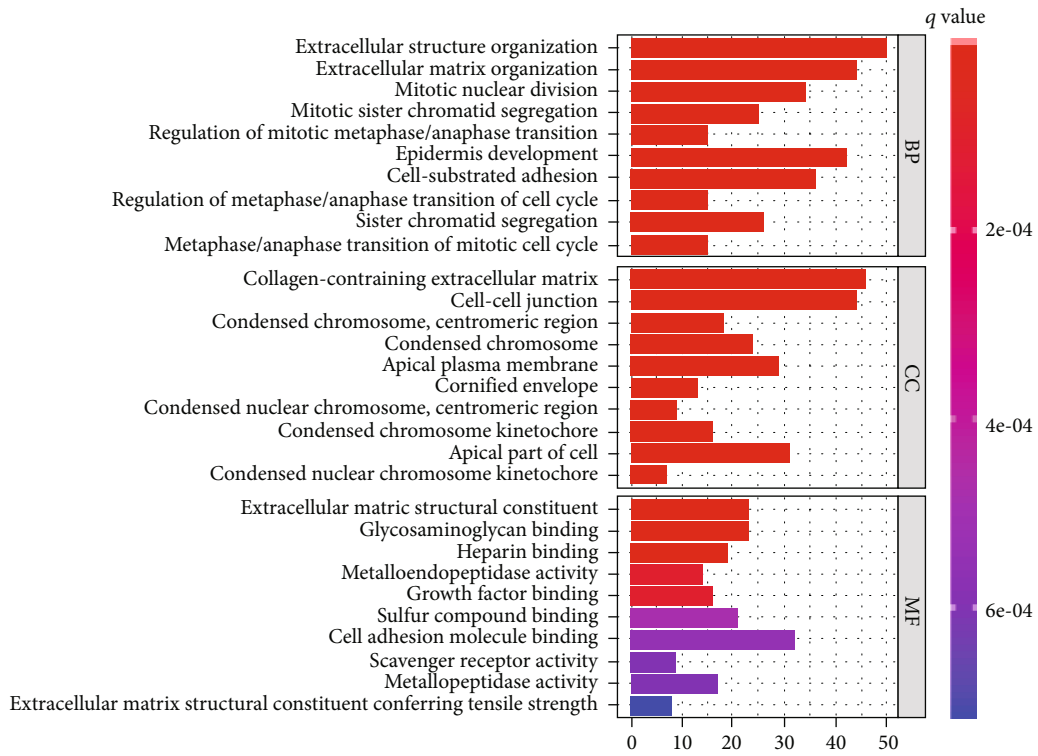


(a)

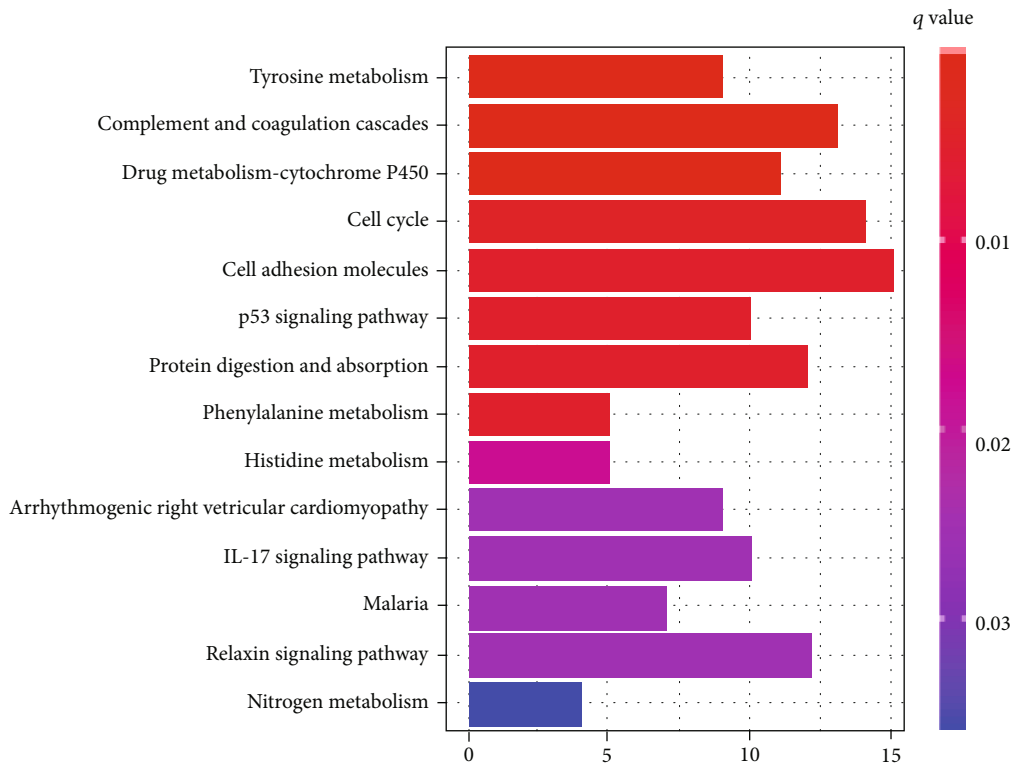


(b)

FIGURE 2: Continued.



(c)



(d)

FIGURE 2: Continued.

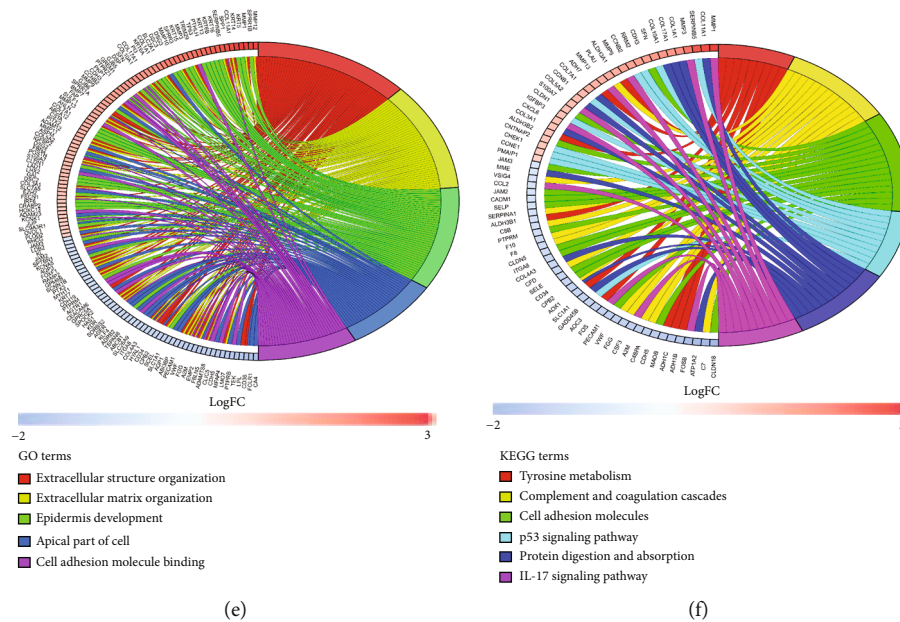


FIGURE 2: Identification and functional enrichment analyses of robust DEGs for LUSC. (a) Volcano plots of the DEGs profile in AD from 6 microarray datasets (GSE1987, GSE2088, GSE8569, GSE21933, GSE33479, and GSE33532), red dots represent the upregulated genes, and green dots represent the downregulated genes; (b) heatmap of the top 20 DEGs identified using the RRA method. Red and green dots represent the upregulated and downregulated genes, respectively; (c) GO enrichment analyses of robust DEGs in three parts: biological process (BP), cellular component (CC), and molecular function (MF); (d) KEGG enrichment analyses of robust DEGs; (e) GO terms enrichment analysis of overlapping robust DEGs; (f) KEGG pathway enrichment analysis of overlapping robust DEGs.

Finally, the values missing and normal samples were removed from a matrix with 501 samples and achieved 495 patients to perform survival analysis using the R packages “survminer” and “survival” according to the best cutoff value. p values < 0.05 were recognized as a statistically significant difference.

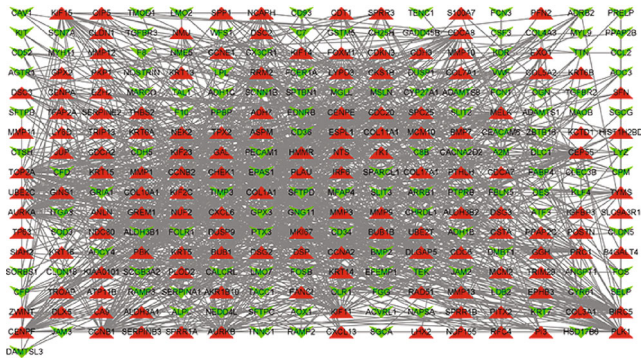
3. Results

3.1. Identification and Functional Enrichment Analyses of Robust DEGs for LUSC. To discover diagnostically and therapeutically novel biomarkers for LUSC, integrated bioinformatics strategies were employed to determine the biological characteristics of robust DEGs from six microarray datasets (Figure 1). First, we investigated the characteristics and DEGs in each dataset and demonstrated that the data quality of six datasets is suitable for RRA analysis. The six datasets included 203 samples consisting of 73 pericarcinomatous tissues and 130 LUSC tissues (Table 1). The distribution of upregulated (red) or downregulated (green) genes is exhibited in volcano plots (Figure 2(a)). We then integrated the six datasets with minimal bias using the “RobustRankAggregation” package to acquire the robust DEGs between normal and tumor tissues. A total of 513 robust DEGs were identified including 220 upregulated and 293 downregulated genes (Supplementary Materials, Table S1). The cutoff criteria for the DEGs were $|\log_2 \text{fold change (FC)}| > 1$ and $\text{FDR} <$

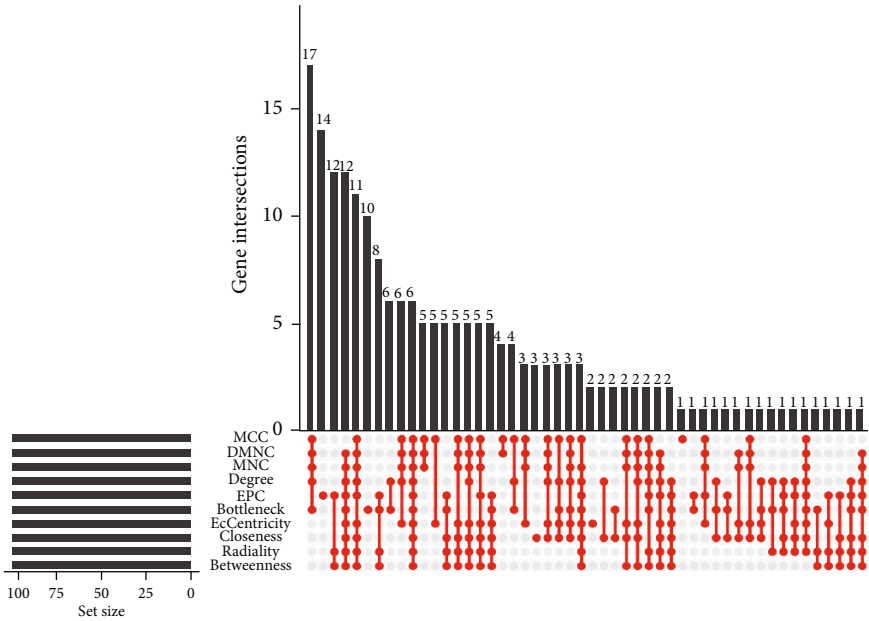
0.05. The top 20 upregulated and downregulated robust DEGs are presented in a visualized heatmap (Figure 2(b)).

In order to explore the function of the robust DEGs in LUSC development, we carried out GO and KEGG pathway enrichment analyses. Nine hundred and one GO terms were found with q value < 0.05 (Supplementary Materials, Table S2). These robust DEGs were most significantly enriched in the extracellular matrix and structure organization, epidermis development, cell adhesion molecule binding, and apical part of cells (Figure 2(c)). The KEGG pathway enrichment analysis showed that tyrosine metabolism, complement and coagulation cascades, cell adhesion molecules, p53 signaling pathway, and IL-17 signaling pathway were the most significantly affected phases in LUSC (Figure 2(d); Supplementary Materials, Table S3). Moreover, SPP1, COL7A1, JUP, and GAL stood out in multiple GO and KEGG terms by overlapping robust DEGs analyses (Figures 2(e) and 2(f)), suggesting that they play important roles in LUSC development.

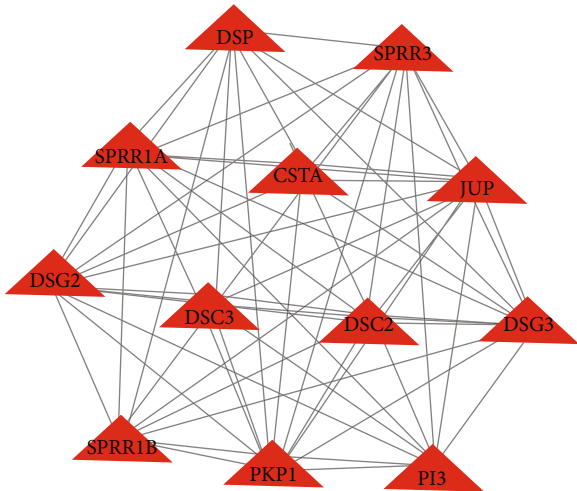
3.2. Identification of Hub Genes through Protein-Protein Interaction Network and Topological Analysis Algorithm Analyses. To explore the hub genes in LUSC progression, we first constructed a visualized PPI network of the robust DEGs with a confidence > 0.9 and dismissed the disconnected nodes, which possessed 273 nodes and 1223 edges (including 143 upregulated and 130 downregulated genes, Figure 3(a)). Then, we evaluated the scores of each node in



(a)



(b)



(c)

FIGURE 3: Continued.

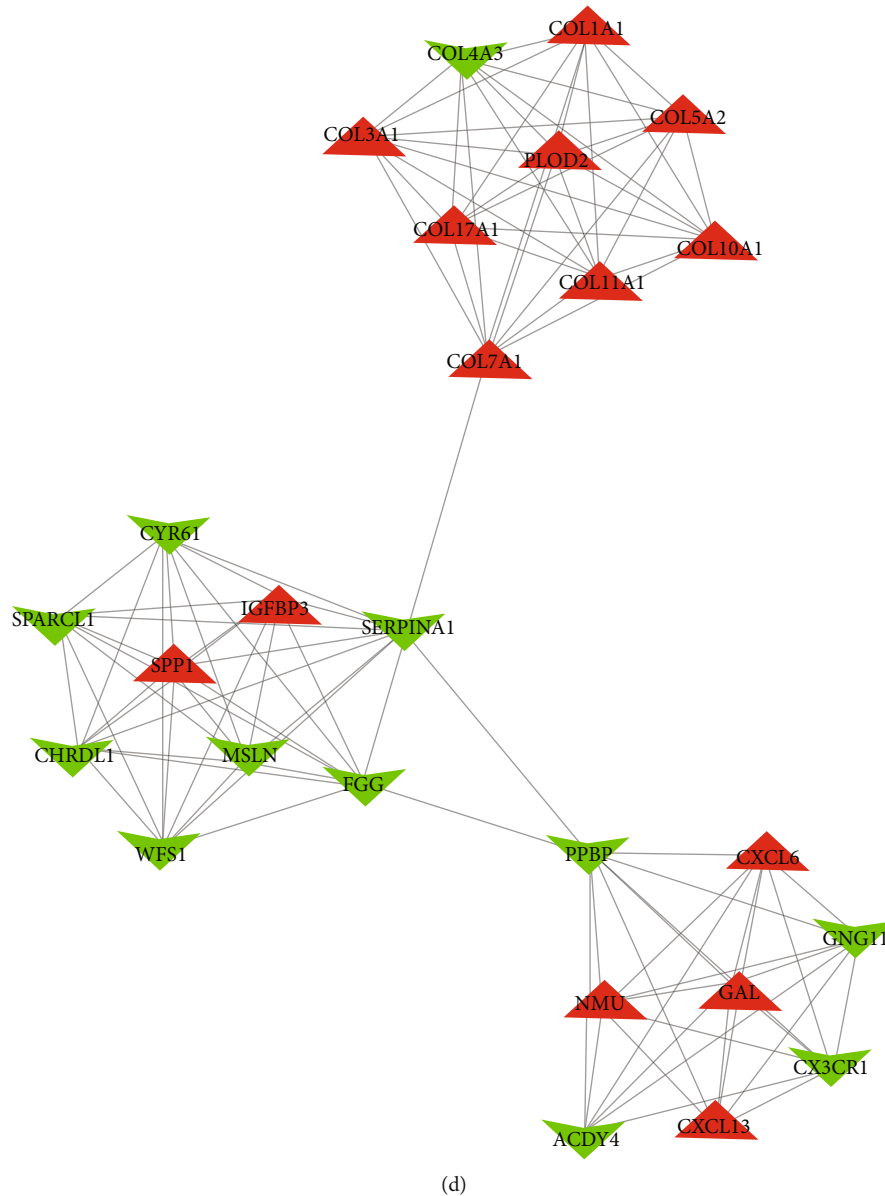


FIGURE 3: Identification of hub genes through PPI network and key modules. (a) The whole PPI network of the robust DEGs. Red and green nodes represent upregulated and downregulated genes, respectively; (b) Module 1 from the whole PPI network containing the JUP hub gene; (c) Module 2 from the whole PPI network including SPP1, COL7A1, GAL, JUP, MSLN, and CHRDL1 hub genes; (d) identifying the hub genes using 10 algorithms including MCC, DMNC, MNC, Degree, EPC, BottleNeck, EcCentricity, Closeness, Radiality, and Betweenness.

the PPI networks with 10 topological analysis algorithms, including MCC, DMNC, MNC, Degree, EPC, BottleNeck, EcCentricity, Closeness, Radiality, and Betweenness, utilizing the Cytoscape plugin cytoHubba (Figure 3(b)). Six hub genes, SPP1, COL7A1, GAL, JUP, MSLN, and CHRDL1, came out of the top 100 genes in the whole network and were considered the hub genes in LUSC progression. The functions and information of the six hub genes are described in Table 2. In addition, we employed the MCODE plugin to identify two key modules associated with six hub genes from the whole PPI network. One module (MCODE score = 12) possessed 12 upregulated genes and 66 edges, containing

the JUP hub gene, which is mainly enriched in cornification, epidermal cell differentiation, and epidermis development pathways (Figure 3(c)). Another module has (MCODE score = 8.24) 26 nodes and 103 edges, including SPP1, COL7A1, GAL, MSLN, and CHRDL1 hub genes, enriched chiefly in the extracellular matrix organization, endoplasmic reticulum lumen, and chemokine-mediated signaling pathways (Figure 3(d)). These data indicated that the six hub genes take part in the regulation of LUSC progress.

3.3. Correlation between Hub Genes and Tumor-Infiltrating Monocytes. Since the tumor immune microenvironment

TABLE 2: The information and functions of the 6 hub genes.

Gene	Full name	Synonyms	Function
SPP1	Osteopontin	BNSP, OPN, PSEC0156	Major noncollagenous bone protein, cell-matrix interaction, cytokine enhancer
COL7A1	Collagen alpha-1(VII) chain		Squamous epithelial basement membrane protein, epithelial basement membrane organization and adherence
JUP	Junction plakoglobin	CTNNG, DP3	Junctional plaque protein, alpha-catenin binding, cadherin binding, cell adhesion molecule binding
GAL	Galanin peptides	GAL1, GALN, GLNN	Endocrine hormone of the central and peripheral nervous systems
MSLN	Mesothelin	MPF	Membrane-anchored forms may play a role in cellular adhesion, cell adhesion, cell-matrix adhesion
CHRDL1	Chordin-like protein 1	NRLN1	Negative regulation of BMP signaling pathway, cell differentiation

plays an important role in tumorigenesis [9, 30], we identified the abundance ratios (p values of <0.05) of 22 types of immune cells in normal and LUSC samples with the CIBERSORT algorithm. The results from six GEO datasets showed that the fractions for T-cell CD8, T-cell CD4 memory resting, NK cells resting, monocytes, mast cells resting, and neutrophils in LUSC tissues were substantially lower than those in normal controls, while the fractions for T-cell follicular helper and macrophages M0 and M1 in LUSC tissues were higher compared with those in normal controls (Figures 4(a) and 4(b)). To further verify these above findings, we performed RNA sequencing for five paired LUSC and normal tissue samples and calculated the abundance ratios of immune cell infiltration. The data also demonstrated that the fractions of monocytes were significantly lower in LUSC tissues than that in normal controls (Figures 4(c) and 4(d)). The immune infiltrating abundance of monocytes was further determined using the six GEO datasets and RNA sequencing data. As shown in Figure 4(e), there were significant differences between LUSC patients and controls. Additionally, the results of PCA, based on the CIBERSORT-calculated results from six GEO datasets and RNA sequencing data, indicated the substantial individual difference between normal and LUSC patients (Figure 4(f)).

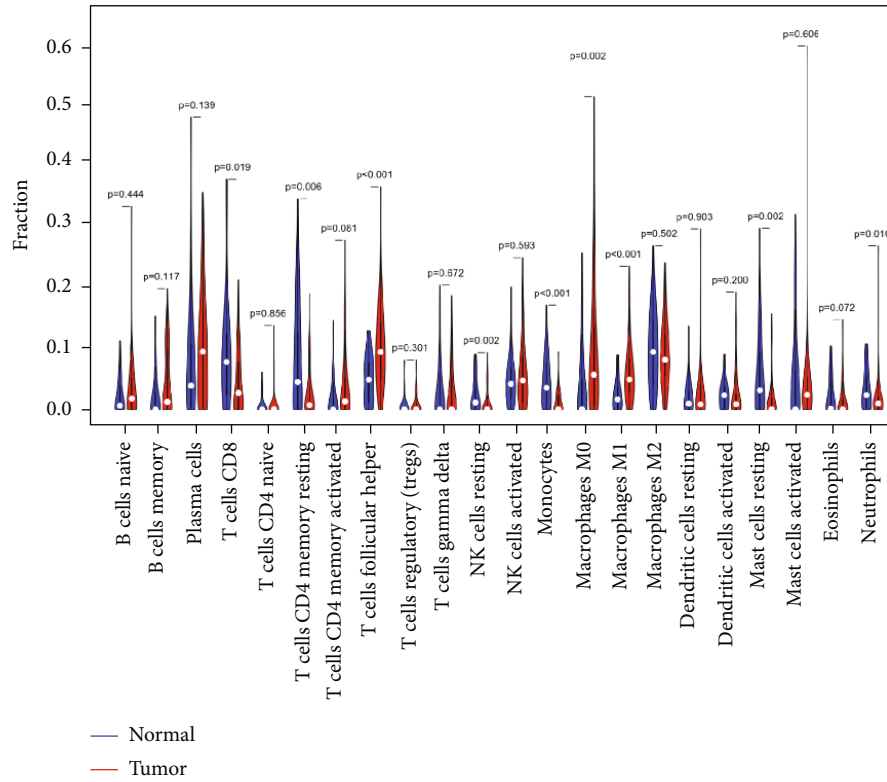
We subsequently verified the expression profiles of six hub genes in TCGA and GTEx datasets, as well as RNA sequencing data from clinical samples. The expression results of six hub genes, SPP1, COL7A1, GAL, and JUP were upregulated, and those of MSLN and CHRDL1 were downregulated in LUSC compared to controls (Figure 5), similar to the results from the integrated six GEO datasets (Table S1). Furthermore, we employed Timer 2.0 to explore the correlation between hub genes and tumor-infiltrating monocytes. The results indicated that the upregulated hub genes COL7A1 and JUP were negatively correlated with tumor-infiltrating monocytes, while the downregulated hub genes MSLN and CHRDL1 were positively correlated with tumor-infiltrating monocytes, with a p value < 0.05 (Figure 6).

3.4. Patient Survival Analysis of Hub Genes. To investigate if there was any prognostic value for the identified hub genes, we analyzed the clinical cases and gene expression profile data of 595 LUSC patients in TCGA database. The patients were divided into two groups according to the optimal cutoff value of hub genes to determine their overall survival rate (Supplementary Materials, Table S4). We observed that the overall survival rates of LUSC patients were significantly associated with the expression of SPP1, COL7A1, GAL, MSLN, and CHRDL1, but not with JUP (Figure 7). Combined with correlation analysis results, we found that COL7A1, MSLN, and CHRDL1 were considered potential prognostic biomarkers associated with LUSC-infiltrated monocytes, indicating the important roles of these hub genes in the pathogenesis of LUSC.

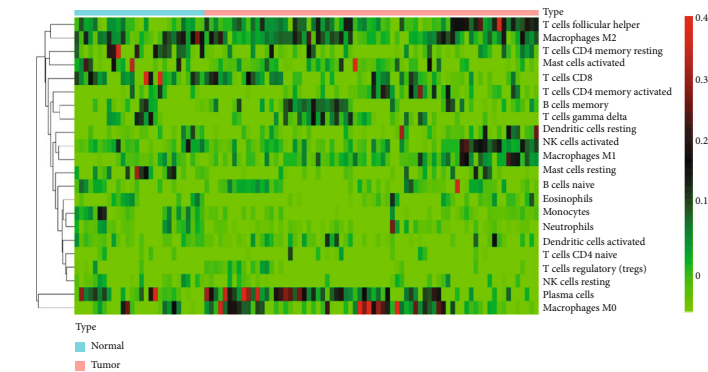
4. Discussion

Many LUSC datasets and GEO databases are not being utilized very well because of small sizes and potential bias. In this study, we integrated 6 microarray datasets with low errors and noise to overcome the above shortcomings by using the RRA method and identified 513 robust DEGs including 220 upregulated and 293 downregulated genes. GO term and KEGG pathway analyses showed that these robust DEGs were significantly involved in the extracellular matrix and structure organization, epidermis development, cell adhesion molecule binding, apical part of cell, p53 signaling pathway, and IL-17 signaling pathway to affect phases in LUSC [31, 32]. Three hub genes of COL7A1, MSLN, and CHRDL1 were considered potential prognostic biomarkers associated with LUSC-infiltrating monocytes and played important roles in the pathogenesis of LUSC.

Immunotherapy undoubtedly is a revolutionized treatment for cancer [33]. Nonetheless, the performance in prolonged survival depends on the individual immune microenvironment [34]. Tumor-infiltrating monocytes are critical regulators in the tumor immune microenvironment, modulating tumor growth and metastasis [35, 36]. The



(a)



(b)

FIGURE 4: Continued.

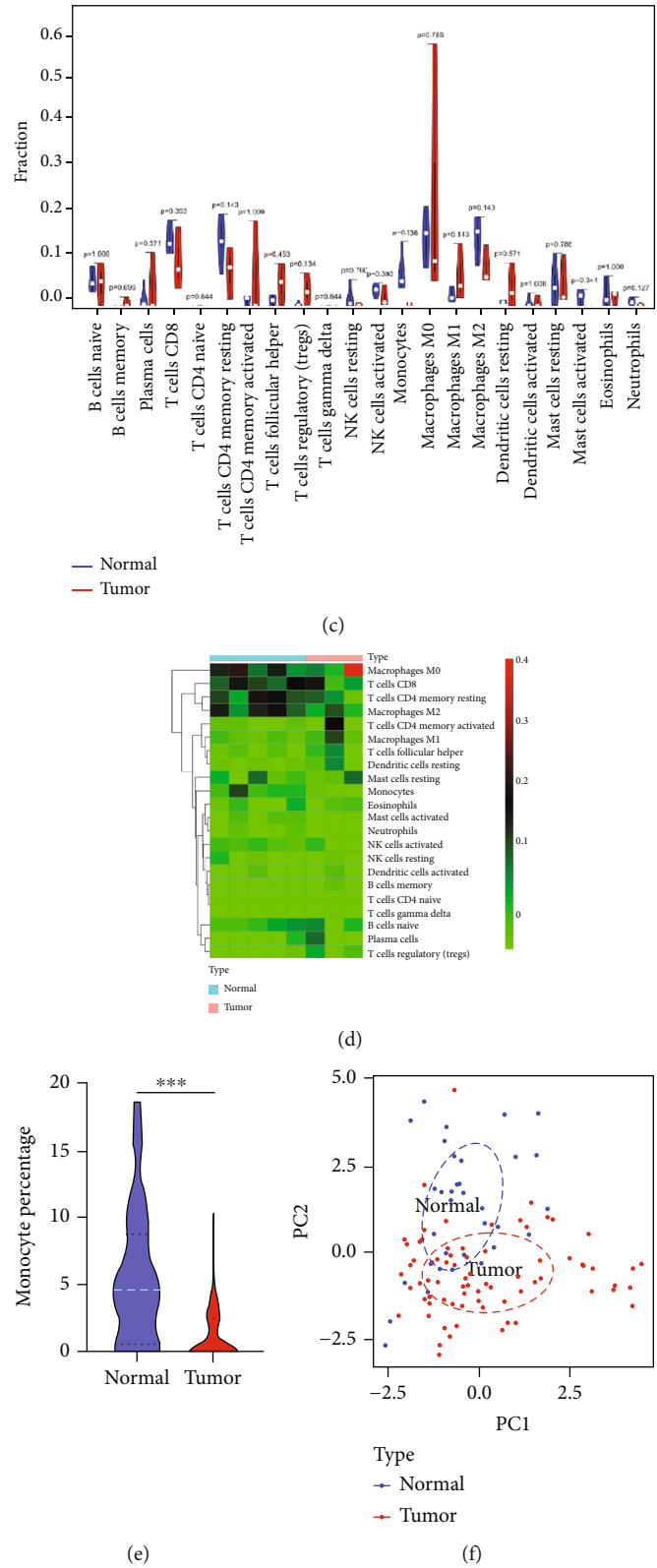


FIGURE 4: Immune cell infiltration analysis. (a) A violin plot of the differentially infiltrated immune cells in normal and LUSC lung tissues from six GEO datasets; (b) the differences in immune cell infiltration between normal and LUSC lung tissues from six GEO datasets shown in a heatmap; (c) a violin plot of the differentially infiltrated immune cells in normal and LUSC lung tissues from RNA sequencing data; (d) the differences of immune cell infiltration between normal and LUSC lung tissues from RNA sequencing data shown in a heatmap; (e) the boxplot of monocytes between LUSC patients and controls in six GEO datasets and RNA sequencing data (***) represents $p < 0.001$; (f) principal component analysis for the normal and LUSC lung tissues.

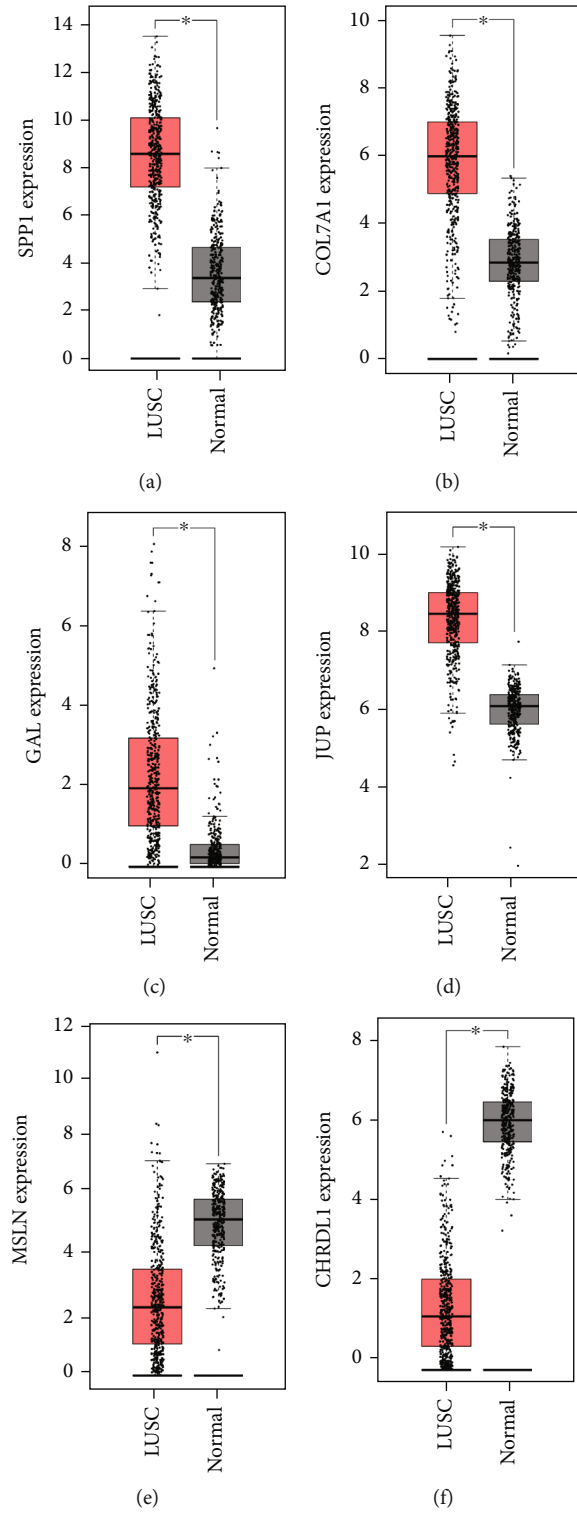


FIGURE 5: Continued.

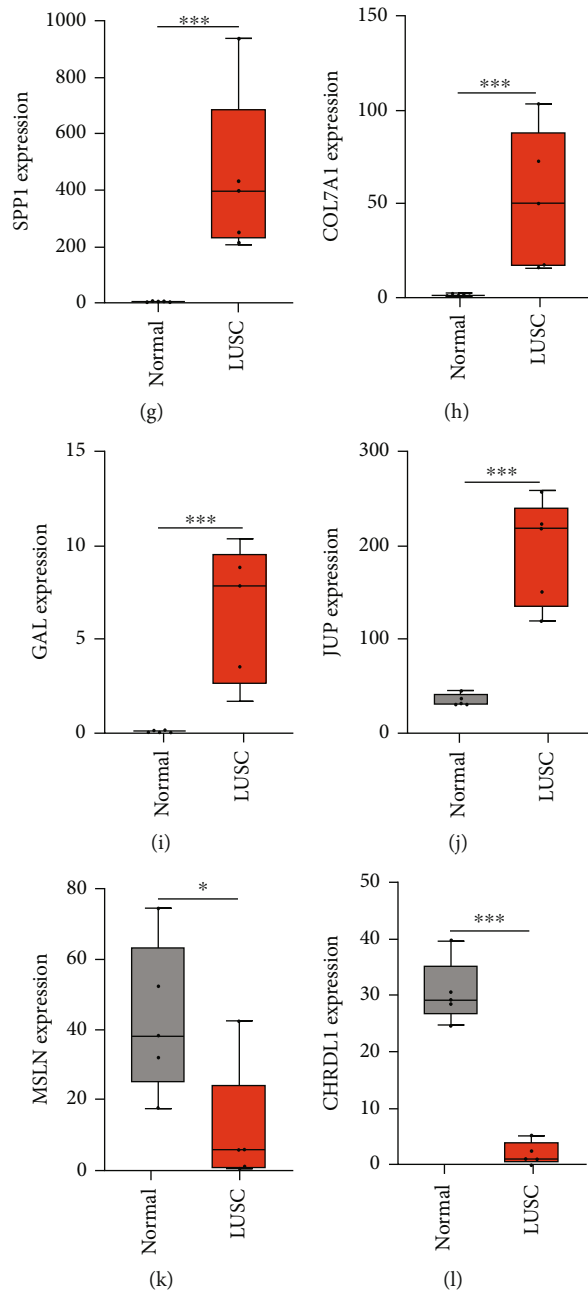


FIGURE 5: The expression analysis of hub genes. (a–f) The expression analysis of hub genes through GEPIA2; (g–l) the expression analysis of hub genes by RNA sequencing.

activity depends on the plasticity of monocytes in response to the stimuli of TME. The expression of hub genes in tumorigenesis determines the profile of immune cell infiltration in TME [37]. Zhang et al. found that four hub genes, LAPTM5, C1QC, CSF1R, and SLCO2B1, could regulate immune cell infiltration of TME in LUSC [19]. Based on the results of immune cell infiltration analysis from six GEO datasets and RNA sequencing data of our clinical sample, we found that the fractions of tumor-infiltrating monocytes were significantly lower in LUSC tissues than that in normal controls. The expression of three hub genes,

COL7A1, MSLN, and CHRDL1, was significantly correlated with tumor-infiltrating monocytes.

Collagen alpha-1(VII) chain encoded by the COL7A1 gene is a kind of stratified squamous epithelial basement membrane protein [38]. It can form anchoring fibrils and interact with extracellular matrix (ECM) proteins which may contribute to epithelial basement membrane organization and adherence [39]. The COL7A1 expression is correlated with tumor invasion and prognosis in esophageal squamous cell carcinoma [40]. The expression of COL7A1 showed a significant prognostic value for OS and distant

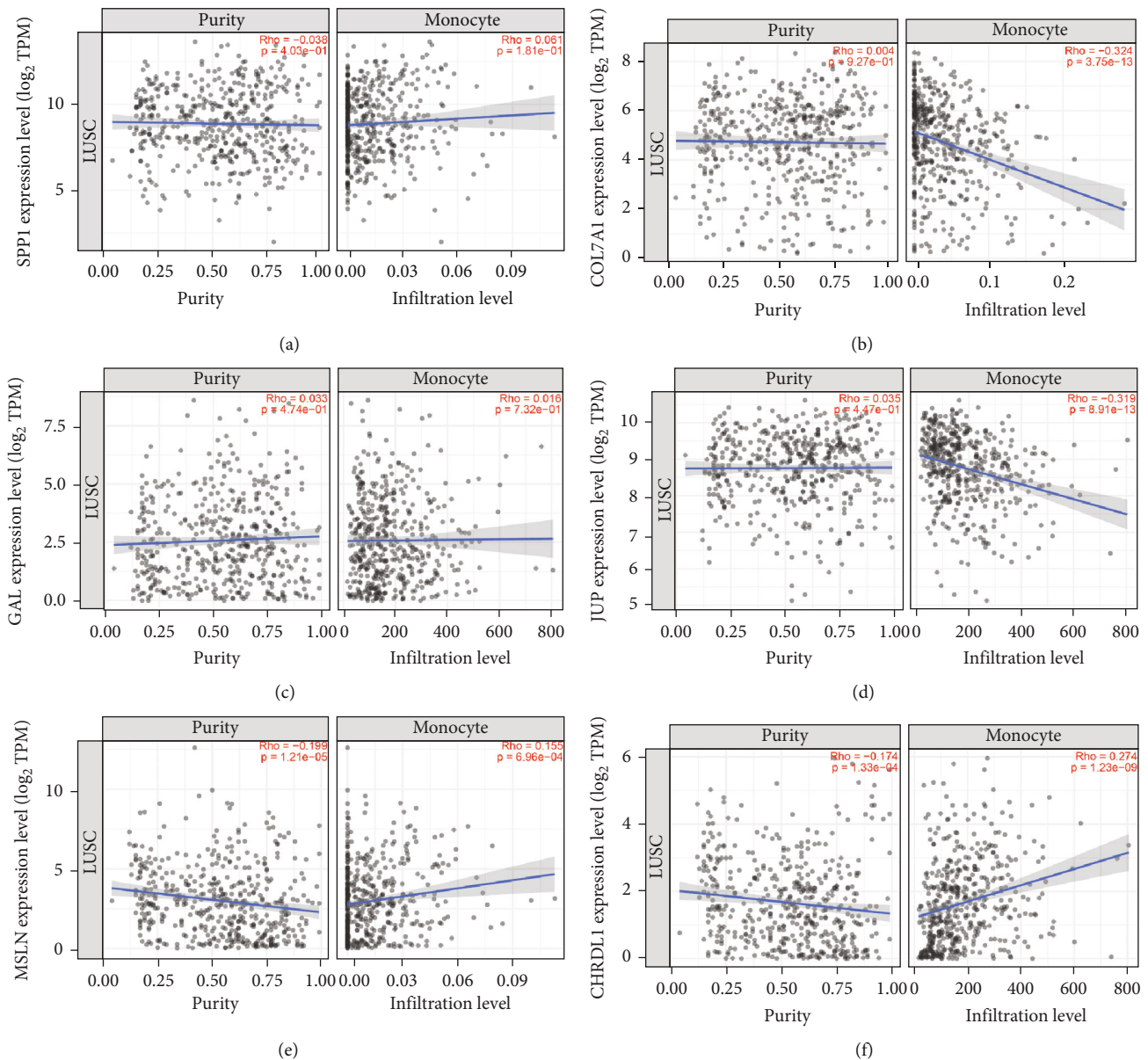
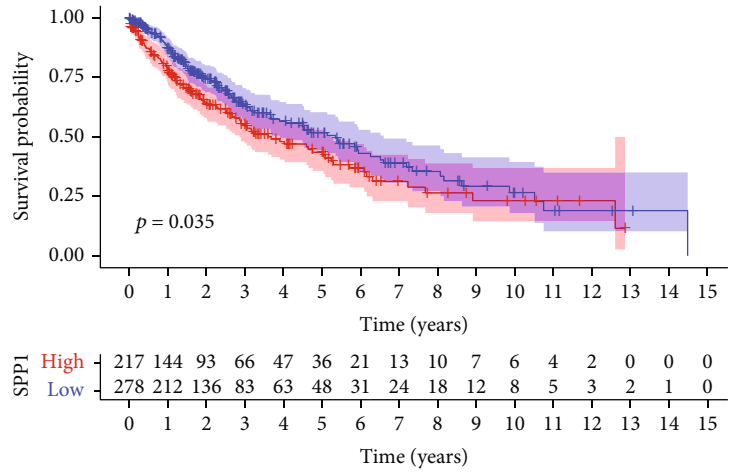


FIGURE 6: The correlation analysis with infiltrating monocytes for hub genes. (a–f) The correlation analysis of hub genes with infiltrating monocytes via Timer 2.0.

metastasis in gastric cancer [41]. We found that the expression of COL7A1 was negatively correlated with tumor-infiltrating monocytes which obviously may be considered an indicator of the overall survival rates of LUSC patients. Mesothelin (encoded by the MSLN gene) involves in cell adhesion and cell-matrix adhesion as membrane-anchored forms, which is also a target of CAR-T cells for treating gastric cancer and ovarian cancer triple-negative breast cancer [42–44]. In gastric cancer, MSLN-CAR-NK cells show strong antitumor activity [45]. We also observed that MSLN was a good prognostic biomarker, and its expression was positively correlated with tumor-infiltrating monocytes, which may possess the role of MSLN-specific CAR monocytes in LUSC. Chordin-like 1 (another hub gene of

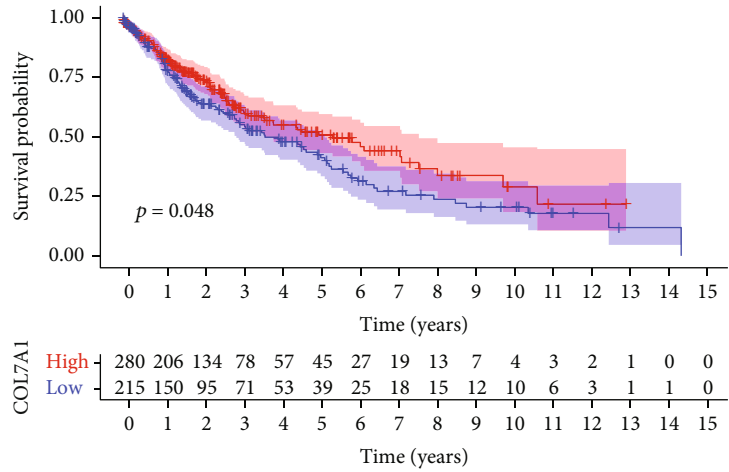
CHRDL1), an inhibitor of BMP, could participate in tumorigenesis [46]. It has been recently reported that CHRDL1 may regulate immune cell infiltration to facilitate immunotherapy. It might be a novel prognostic biomarker and therapy target in LUAD [47]. However, there was no evidence that it has been investigated in LUSC until now. Our study demonstrated that CHRDL1 was negatively correlated with tumor-infiltrating monocytes as well as may be a novel prognostic biomarker and therapy target in LUSC. These facts suggest that three hub genes of COL7A1, MSLN, and CHRDL1 are important potential prognostic biomarkers and therapy targets in LUSC.

There were two limitations in this study. External dataset evaluation is needed to enhance the accuracy of diagnostic



SPP1
+ High
+ Low

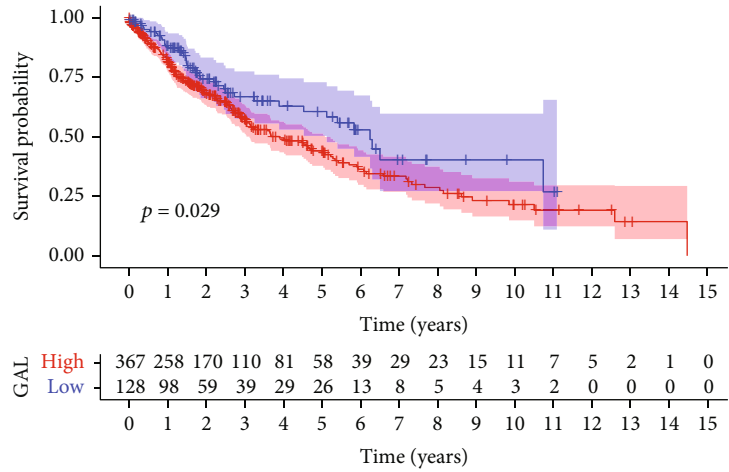
(a)



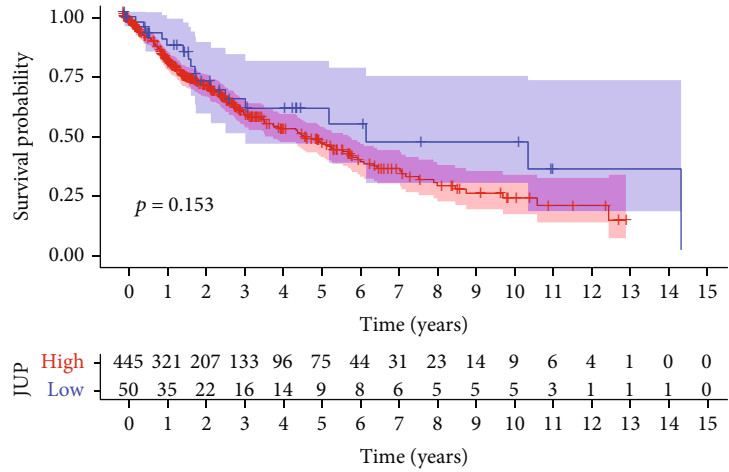
COL7A1
+ High
+ Low

(b)

FIGURE 7: Continued.



(c)



(d)

FIGURE 7: Continued.

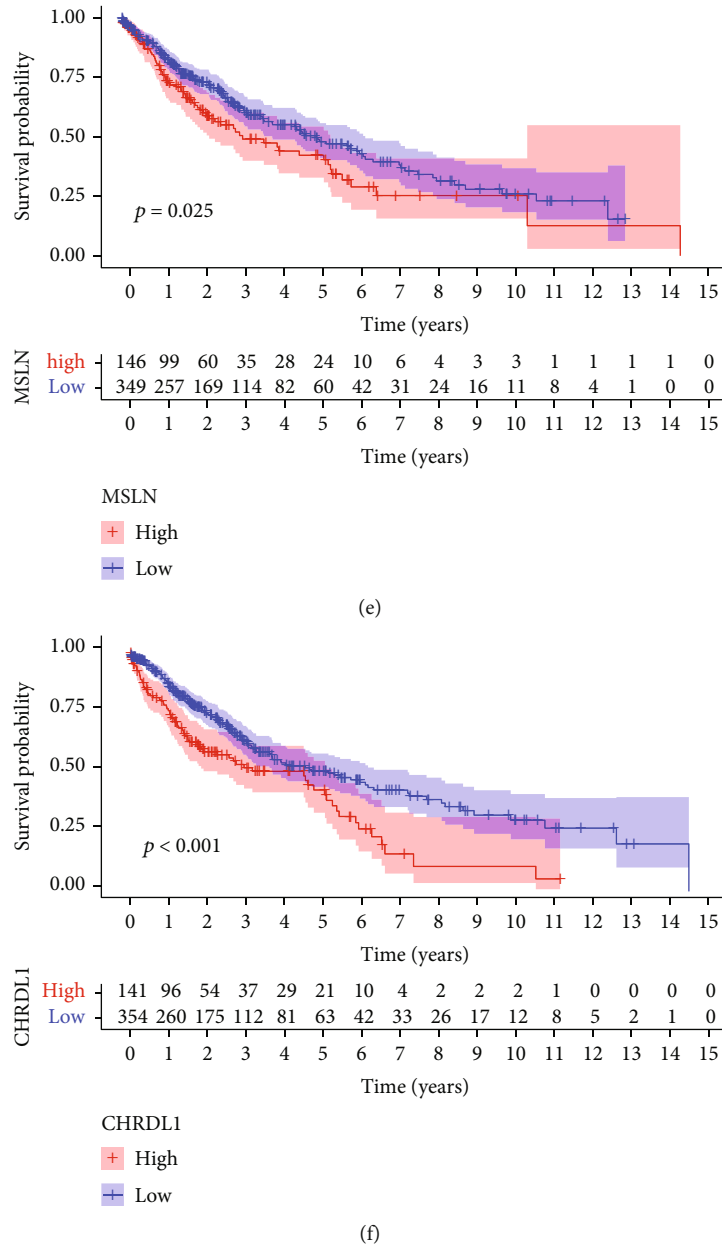


FIGURE 7: Survival analysis of hub genes in LUSC. The overall survival curves of USC patients based on the expression of SPP1 (a), COL7A1 (b), GAL (c), JUP (d), MSLN (e), and CHRDL1 (f), respectively.

genes. Secondly, laboratory and clinical validation is important and needs to be investigated for hub genes and infiltrating monocytes in the future.

5. Conclusion

In the study, we conduct RRA algorithms to identify 513 robust DEGs, including 220 upregulated and 293 downregulated genes, from 203 samples in six LUSC-GEO datasets with low errors and noise. Three hub genes, COL7A1, MSLN, and CHRDL1, were identified from 513 robust DEGs using a series of bioinformatics methods, which were significantly correlated with tumor-infiltrating monocytes as well as effective indicators of prognostic outcomes for LUSC

patients. The research of these gene markers can improve to well understand the molecular mechanisms of lung squamous cell carcinoma development.

Abbreviations

- LUSC: Lung squamous cell carcinoma
- NSCLC: Non-small-cell lung cancer
- TME: Tumor microenvironment
- TICs: Tumor-infiltrating immune cells
- RRA: Robust rank aggregation
- DEGs: Differentially expressed genes
- GEO: Gene Expression Omnibus
- GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes
 PPI: Protein-protein interaction
 BP: Biological process
 CC: Cellular component
 MF: Molecular function.

Data Availability

The datasets of this article were downloaded from TCGA database and the GEO database.

Conflicts of Interest

The authors confirm that there are no conflicts of interest.

Authors' Contributions

HLL designed the study, reviewed the publications, and wrote this manuscript. BY, YLC, and JP collected the information and processed samples. YL, CGL, and ZFZ participated in the study design, data collection, analysis of data, and preparation of the manuscript. TTQ analyzed and drafted the manuscript. All authors read and approved the final version of the manuscript. Hailin Liu, Bo Yan, and Yulong Chen contributed equally to this work.

Acknowledgments

Thanks are due to the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (2017KJ201, Tingting Qin).

Supplementary Materials

Table S1: the gene lists of robust differentially expressed genes. A total of 513 robust DEGs were identified including 220 upregulated and 293 downregulated genes. Table S2: the lists of GO terms for robust differentially expressed genes. Nine hundred and one GO terms were found with q value < 0.05 . Table S3: the lists of KEGG terms for robust differentially expressed genes. The KEGG pathway enrichment analysis. Table S4: the cutoff value and the p value of each hub gene for the overall survival of patients. The overall survival rates of LUSC patients were significantly associated with the expression of SPP1, COL7A1, GAL, MSLN, and CHRDL1, but not with JUP. (*Supplementary Materials*)

References

- [1] H. Sung, J. Ferlay, R. L. Siegel et al., "GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *Ca: A Cancer Journal For Clinicians* 2021, vol. 71, no. 3, pp. 209–249, 2021.
- [2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: a Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, 2022.
- [3] P. Torres-Ayuso, E. An, K. M. Nyswaner et al., "TNIK is a therapeutic target in lung squamous cell carcinoma and regulates FAK activation through merlin," *Cancer Discovery*, vol. 11, no. 6, pp. 1411–1423, 2021.
- [4] F. Barbone, M. Bovenzi, F. Cavallieri, and G. Stanta, "Cigarette smoking and histologic type of lung cancer in men," *Chest*, vol. 112, no. 6, pp. 1474–1479, 1997.
- [5] J. R. Molina, P. Yang, S. D. Cassivi, S. E. Schild, and A. A. Adjei, "Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship," *Mayo Clinic Proceedings*, vol. 83, no. 5, pp. 584–594, 2008.
- [6] D. A. Siegel, M. E. O'Neil, T. B. Richards, N. F. Dowling, and H. K. Weir, "Prostate cancer incidence and survival, by stage and race/ethnicity - United States, 2001-2017," *MMWR. Morbidity and Mortality Weekly Report*, vol. 69, no. 41, pp. 1473–1480, 2020.
- [7] Y. Jiang, H. Yuan, Z. Li et al., "Global pattern and trends of colorectal cancer survival: a systematic review of population-based registration data," *Cancer Biology & Medicine*, vol. 18, 2021.
- [8] L. Mangone, P. Mancuso, I. Bisceglia et al., "Five-year relative survival by stage of breast and colon cancers in Italy," *Tumori*, vol. 107, no. 4, pp. 318–324, 2021.
- [9] X. Lei, Y. Lei, J. K. Li et al., "Immune cells within the tumor microenvironment: biological functions and roles in cancer immunotherapy," *Cancer Letters*, vol. 470, pp. 126–133, 2020.
- [10] A. K. Ojlert, A. R. Halvorsen, D. Nebdal et al., "The immune microenvironment in non-small cell lung cancer is predictive of prognosis after surgery," *Molecular Oncology*, vol. 13, no. 5, pp. 1166–1179, 2019.
- [11] Y. Yang, G. Wu, Q. Li et al., "Angiogenesis-related immune signatures correlate with prognosis, tumor microenvironment, and therapeutic sensitivity in hepatocellular carcinoma," *Frontiers in Molecular Biosciences*, vol. 8, article 690206, 2021.
- [12] R. Huang, M. Mao, Y. Lu, Q. Yu, and L. Liao, "A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment," *Aging (Albany NY)*, vol. 12, no. 8, pp. 6966–6980, 2020.
- [13] J. Wang, W. Shi, Y. Miao, J. Gan, Q. Guan, and J. Ran, "Evaluation of tumor microenvironmental immune regulation and prognostic in lung adenocarcinoma from the perspective of purinergic receptor P2Y13," *Bioengineered*, vol. 12, no. 1, pp. 6286–6304, 2021.
- [14] M. Farha, N. K. Jairath, T. S. Lawrence, and I. El Naqa, "Characterization of the tumor immune microenvironment identifies M0 macrophage-enriched cluster as a poor prognostic factor in hepatocellular carcinoma," *JCO Clinical Cancer Informatics*, vol. 4, pp. 1002–1013, 2020.
- [15] B. Liu, X. Ma, and W. Ha, "Identification of potential prognostic biomarkers associated with macrophage M2 infiltration in gastric cancer," *Frontiers in Genetics*, vol. 12, article 827444, 2021.
- [16] C. E. Olingy, H. Q. Dinh, and C. C. Hedrick, "Monocyte heterogeneity and functions in cancer," *Journal of Leukocyte Biology*, vol. 106, no. 2, pp. 309–322, 2019.
- [17] R. N. Hanna, C. Cekic, D. Sag et al., "Patrolling monocytes control tumor metastasis to the lung," *Science*, vol. 350, no. 6263, pp. 985–990, 2015.
- [18] Y. Hai, N. Chen, W. Wu et al., "High postoperative monocyte indicates inferior clinicopathological characteristics and worse prognosis in lung adenocarcinoma or squamous cell carcinoma after lobectomy," *BMC Cancer*, vol. 18, no. 1, p. 1011, 2018.
- [19] T. Zhang, H. Yang, B. Sun, and F. Yao, "Four hub genes regulate tumor infiltration by immune cells, antitumor immunity

- in the tumor microenvironment, and survival outcomes in lung squamous cell carcinoma patients,” *Aging (Albany NY)*, vol. 13, no. 3, pp. 3819–3842, 2021.
- [20] E. Clough and T. Barrett, “The Gene Expression Omnibus database,” *Methods in Molecular Biology*, vol. 1418, pp. 93–110, 2016.
- [21] M. Gao, W. Kong, Z. Huang, and Z. Xie, “Identification of key genes related to lung squamous cell carcinoma using bioinformatics analysis,” *International Journal of Molecular Sciences*, vol. 21, no. 8, p. 2994, 2020.
- [22] Y. Shi, Y. Li, C. Yan, H. Su, and K. Ying, “Identification of key genes and evaluation of clinical outcomes in lung squamous cell carcinoma using integrated bioinformatics analysis,” *Oncology Letters*, vol. 18, no. 6, pp. 5859–5870, 2019.
- [23] R. Kolde, S. Laur, P. Adler, and J. Vilo, “Robust rank aggregation for gene list integration and meta-analysis,” *Bioinformatics*, vol. 28, no. 4, pp. 573–580, 2012.
- [24] Z. Y. Song, F. Chao, Z. Zhuo, Z. Ma, W. Li, and G. Chen, “Identification of hub genes in prostate cancer using robust rank aggregation and weighted gene co-expression network analysis,” *Aging*, vol. 11, no. 13, pp. 4736–4756, 2019.
- [25] B. Zhang, Z. Zhang, H. Hu, and J. T. Dong, “Novel gene signatures predictive of patient recurrence-free survival and castration resistance in prostate cancer,” *Cancers (Basel)*, vol. 13, no. 4, p. 917, 2021.
- [26] P. Shannon, A. Markiel, O. Ozier et al., “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [27] C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, and C. Y. Lin, “cytoHubba: identifying hub objects and sub-networks from complex interactome,” *BMC Systems Biology*, vol. 8, Supplement 4, p. S11, 2014.
- [28] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, and A. A. Alizadeh, “Profiling tumor infiltrating immune cells with CIBERSORT,” *Methods in Molecular Biology*, vol. 1711, pp. 243–259, 2018.
- [29] T. Li, J. Fu, Z. Zeng et al., “TIMER2.0 for analysis of tumor-infiltrating immune cells,” *Nucleic Acids Research*, vol. 48, pp. W509–W514, 2020.
- [30] M. Binnewies, E. W. Roberts, K. Kersten et al., “Understanding the tumor immune microenvironment (TIME) for effective therapy,” *Nature Medicine*, vol. 24, no. 5, pp. 541–550, 2018.
- [31] V. K. de Sa, L. Carvalho, A. Gomes et al., “Role of the extracellular matrix in variations of invasive pathways in lung cancers,” *Brazilian Journal of Medical and Biological Research*, vol. 46, no. 1, pp. 21–31, 2013.
- [32] I. Shi, N. Hashemi Sadraei, Z. H. Duan, and T. Shi, “Aberrant signaling pathways in squamous cell lung carcinoma,” *Cancer Informatics*, vol. 10, pp. 273–285, 2011.
- [33] S. Ugel, S. Cane, F. De Sanctis, and V. Bronte, “Monocytes in the tumor microenvironment,” *Annual Review of Pathology*, vol. 16, no. 1, pp. 93–122, 2021.
- [34] R. S. Riley, C. H. June, R. Langer, and M. J. Mitchell, “Delivery technologies for cancer immunotherapy,” *Nature Reviews. Drug Discovery*, vol. 18, no. 3, pp. 175–196, 2019.
- [35] S. Y. Lim, A. E. Yuzhalin, A. N. Gordon-Weeks, and R. J. Muschel, “Tumor-infiltrating monocytes/macrophages promote tumor invasion and migration by upregulating S100A8 and S100A9 expression in cancer cells,” *Oncogene*, vol. 35, no. 44, pp. 5735–5745, 2016.
- [36] S. W. L. Lee, G. Adriani, R. D. Kamm, and M. R. Gillrie, “Models for monocytic cells in the tumor microenvironment,” *Advances in Experimental Medicine and Biology*, vol. 1224, pp. 87–115, 2020.
- [37] Z. Wu, M. Wang, Q. Liu et al., “Identification of gene expression profiles and immune cell infiltration signatures between low and high tumor mutation burden groups in bladder cancer,” *International Journal of Medical Sciences*, vol. 17, no. 1, pp. 89–96, 2020.
- [38] A. Nystrom, L. Bruckner-Tuderman, and D. Kiritsi, “Dystrophic epidermolysis bullosa: secondary disease mechanisms and disease modifiers,” *Frontiers in Genetics*, vol. 12, article 737272, 2021.
- [39] K. Gelse, E. Poschl, and T. Aigner, “Collagens—structure, function, and biosynthesis,” *Advanced Drug Delivery Reviews*, vol. 55, no. 12, pp. 1531–1546, 2003.
- [40] Y. Kita, K. Mimori, F. Tanaka et al., “Clinical significance of *_LAMB3_* and *_COL7A1_* mRNA in esophageal squamous cell carcinoma,” *European Journal of Surgical Oncology*, vol. 35, no. 1, pp. 52–58, 2009.
- [41] S. E. Oh, M. Y. Oh, J. Y. An et al., “Prognostic value of highly expressed type VII collagen (COL7A1) in patients with gastric cancer,” *Pathology Oncology Research*, vol. 27, p. 1609860, 2021.
- [42] J. Lv, R. Zhao, D. Wu et al., “Mesothelin is a target of chimeric antigen receptor T cells for treating gastric cancer,” *Journal of Hematology & Oncology*, vol. 12, no. 1, p. 18, 2019.
- [43] E. Schoutrop, I. El-Serafi, T. Poiret et al., “Mesothelin-specific CAR T cells target ovarian cancer,” *Cancer Research*, vol. 81, no. 11, pp. 3022–3035, 2021.
- [44] P. Yang, X. Cao, H. Cai et al., “The exosomes derived from CAR-T cell efficiently target mesothelin and reduce triple-negative breast cancer growth,” *Cellular Immunology*, vol. 360, article 104262, 2021.
- [45] B. Cao, M. Liu, J. Huang et al., “Development of mesothelin-specific CAR NK-92 cells for the treatment of gastric cancer,” *International Journal of Biological Sciences*, vol. 17, no. 14, pp. 3850–3861, 2021.
- [46] T. Liu, B. Li, X. F. Zheng et al., “Chordin-like 1 improves osteogenesis of bone marrow mesenchymal stem cells through enhancing BMP4-SMAD pathway,” *Front Endocrinol (Lausanne)*, vol. 10, p. 360, 2019.
- [47] B. Deng, X. Chen, L. Xu et al., “Chordin-like 1 is a novel prognostic biomarker and correlative with immune cell infiltration in lung adenocarcinoma,” *Aging (Albany NY)*, vol. 14, no. 1, pp. 389–409, 2022.