Article

J. Biol. Regul. Homeost. Agents. 2024; 38(6): 4825–4836
https://doi.org/10.23812/j.biol.regul.homeost.agents.20243806.385

Identification of Immune Infiltration Consensus Genes and Their Clinical Value in Early and Advanced Non-Small Cell Lung Carcinoma

Zexin Gu[1], Xiangru Meng[1], Cuicui Li[1], Hanxu Tang[2], Jianing Liu[1], Weiwei Zhao[1,*]

[1]Department of Respiratory Internal Medicine, The Second Affiliated Hospital of Qiqihar Medical University, 161000 Qiqihar, Heilongjiang, China
[2]Medical Examination Section, The Second Affiliated Hospital of Qiqihar Medical University, 161000 Qiqihar, Heilongjiang, China
*Correspondence: zwwqqhe@qmu.edu.cn (Weiwei Zhao)
Submitted: 1 March 2023 Revised: 25 March 2023 Accepted: 15 April 2024 Published: 1 June 2024

Background: Lung cancer stands as the leading cause of cancer-related mortality globally, with non-small cell lung carcinoma (NSCLC) accounting for approximately 85% of all lung cancer cases. Despite advancements in diagnostic techniques and therapeutic interventions, the 5-year survival rate for NSCLC remains low due to the recurrence and dissemination of malignant cells. There is an urgent need to identify novel biomarkers and therapeutic targets to address this challenge. Therefore, this study aims to identify common genes associated with tumor-related immune cells and investigate their potential clinical utility in both early and advanced NSCLC.

Methods: Early-stage and advanced NSCLC expression data, mutation data, and associated medical records were obtained and refined for subsequent examination from The Cancer Genome Atlas (TCGA). Differential expression analysis, gene ontology (GO), transcription factors and pathway enrichment analysis, and gene set enrichment analysis (GSEA) were implemented to discern molecular function and regulatory relationship across differentially expressed genes (DEGs). Single-sample gene set enrichment analysis (ssGSEA) was employed to analyze immune cell abundance. Furthermore, the weighted gene co-expression network analysis (WGCNA) of DEGs was utilized to screen out gene modules related to tumor-associated immune cells in early-stage and advanced NSCLC. This was achieved by the tumor immune estimation resource (TIMER) algorithm to assess immune cell abundance. Subsequently, consensus genes associated with drug sensitivity and pathways activity were analyzed using the Gene Set Cancer Analysis Literate (GSCALite) platform. Notably, we also evaluated the correlation between consensus genes expression and TP53 mutant (TP53mut) and TP53 wild-type (TP53wt). Finally, the KMPlotter online tool was used to evaluate the prognostic implications of consensus genes exhibiting different correlation patterns in NSCLC.

Results: In early and advanced NSCLC, there were 996 (445 upregulations and 551 downregulations) and 822 (398 upregulations and 424 downregulations) DEGs from lung adenocarcinoma (LUAD) versus lung squamous cell carcinoma (LUSC), respectively, following differential expression analysis. In the interferon signal pathway, functional enrichment analysis showed significant enrichment of DEGs. A correlation between immune infiltration and NSCLC was found using ssGSEA. WGCNA analysis revealed a strong association between tumor-immune infiltration characteristics and the blue and turquoise modules. Notably, a total of 27 consensus genes linked to tumor-related immune cells were identified in both early and advanced NSCLC. Furthermore, differential expression patterns were observed for these consensus genes, such as melanoma-associated antigen A 4 (*MAGEA4*) and dynein cytoplasmic 1 intermediate chain 1 (*DYNC1I1*), between TP53 mutant (TP53mut) and TP53 wild-type (TP53wt).

Conclusions: This study revealed the crucial role of immune cell infiltration, especially dendritic cells, in the onset and progression of early and advanced NSCLC, providing potential targets for immune therapy.

Keywords: immune infiltration; NSCLC; TCGA; WGCNA; drug sensitivity; TP53 mutation status

Introduction

Lung cancer is the second most prevalent malignancy worldwide, according to the data from the Global Cancer Statistics 2020. It also exhibits the highest mortality rates [1]. Furthermore, given that breast cancer has the highest incidence, lung metastasis is the main contributor to mortality from breast cancer [2,3]. The incidence of lung cancer is progressively increasing in China annually, with a recent trend showing its development at younger ages [4].

Non-small cell lung carcinoma (NSCLC), which encompasses lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), is the most common subtype of lung cancer [5]. Focusing solely on early-stage NSCLC is insufficient when compared with advanced NSCLC. Research by Guerrera F *et al.* [6] indicated that the prognosis for stage I lung cancer remains unfavorable even after surgical resection, with a 5-year overall survival rate of approximately 70%.

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

Immune checkpoint inhibitors (ICIs) including pembrolizumab, atezolizumab, and durvalumab have been consecutively introduced into clinical practice as potential treatments for NSCLC [7], demonstrating notably effective results. However, clinical practice has revealed instances of therapy failure with ICIs.

Lymphomas are susceptible to immune checkpoint inhibitors that target the PD-L1/PD-1 pathway, attributed to the elevated expression of PD-L1 [8]. The tumor mutation burden (TMB) serves as a significant source of tumor immune antigens. Abnormal proteins are presented on the human leukocyte antigen (HLA) complex and recognized by T cells, thereby triggering an anti-tumor immune response. TMB has emerged as a promising predictive biomarker for the efficacy of immune checkpoint inhibitor therapy [9]. Additionally, the characteristics of the tumor microenvironment and immune checkpoints play a crucial role in determining the response to PD-1/PD-L1 inhibitors [10].

Tumor immune estimation resource (TIMER), enabled by high-throughput sequencing technologies and bioinformatics methods, was employed to assess the distribution of immune cells in cancer [11]. Additionally, gene co-expression modules associated with immune cells can be generated using weighted gene co-expression network analysis (WGCNA), a robust guilt-by-association (GBA) method for constructing co-expression networks [12].

The Cancer Genome Atlas (TCGA) lung cancer dataset was used to compare differential gene expression profiles and associated functional enrichment terms, focusing on immune-related signaling pathways in early and advanced NSCLC, specifically LUAD versus LUSC. Furthermore, immune-related consensus genes were identified using immune cell infiltering analyses and WGCNA. Subsequently, the correlation between consensus genes, drug sensitivity, and signaling pathway activity was investigated. Notably, the presence or absence of a TP53 mutation appeared to influence the expression levels of consensus genes. Patients with lung cancer exhibiting high expression of consensus genes and TP53 mutation generally experienced poorer prognoses.

Immune cell infiltration plays a crucial role in NSCLC. However, the characteristics of immune infiltration in lung cancer subtypes and their association with disease advancement require further investigation. Our study simultaneously analyzed early and advanced NSCLC subtypes, contrasting their immune cell infiltration patterns to delineate similarities and distinctions. This comparative analysis provides insights into the potential application of immunotherapy for lung cancer. By integrating multiple bioinformatics algorithms, we systematically identified common genes related to tumor immunity at the whole-genome level, moving beyond the scope of established immune-related genes. Additionally, the research investigated the associations between consensus genes, drug sensitivity, and important signaling pathway activity, providing clues for discovering potential targeted therapies and predicting treatment efficacy. It also analyzed the effect of TP53 mutation status on consensus gene expression and its correlation with prognosis, laying the foundation for personalized immunotherapy in NSCLC.

Materials and Methods

Data Collection and Preprocessing

Data regarding NSCLC, associated with batch-corrected gene expression profiles, curated clinical data, and unified somatic simple mutation data, were obtained from the PanCanAtlas (https://gdc.cancer.gov/about-data/publications/pancanatlas). Using the HUGO Gene Nomenclature Committee (HGNC) [13] multi-symbol checker tools (https://www.genenames.org/tools/multi-symbol-checker/), all gene names were reannotated to official gene symbols. Quartile normalization was conducted for cross-sample normalization using the "normalizeBetweenArrays" function of the "limma" R package (http://www.rproject.org/). Genes showing zero expression in any sample were excluded, and $\log_2$ (TPM+1) transformation was applied to adjust the data post-analysis. Following the guidelines of the National Comprehensive Cancer Network, we determined the number of non-synonymous and all somatic mutations (tumor mutation burden, TMB) in the coding region for each tumor sample. For this study, we extracted gene expression and associated comprehensive clinical data (**Supplementary Table 1**) from 530 early-stage (stage I, 286 LUAD and 244 LUSC) and 201 advanced-stage (stage III-IV, 110 LUAD and 91 LUSC) NSCLC patients, as described by Shi R *et al*. [14] and Blakely CM *et al*. [15].

Differentially Expressed Genes Analysis

Using the "FactoMineR" (http://www.rproject.org/) and "Factoextra" R packages (http://www.rproject.org/), a principal component analysis (PCA) plot was conducted to evaluate the quality of the transcriptomic data. Differentially expressed genes (DEGs) between LUAD and LUSC were identified employing the "limma" R package, applying established criteria [16] of $|\log_2 FC| > 1$ and adjusted $p < 0.05$. The list of DEGs was then used for downstream analysis, followed by visualization using a volcano plot.

Functional Enrichment Based on Over-Representation Analysis

The background gene set and the Fisher's test are essential components for conducting over-representation analysis (ORA) and the first generation of functional enrichment [17]. Metascape [18] (http://metascape.org/gp/index.html) was implemented to complete the functional annotation for DEGs. Gene/protein functional annotation was facilitated through the gene ontology (GO), encompassing biological processes (BP), cellular components (CC), and molecular functions (MF), as well as the Kyoto Encyclo-

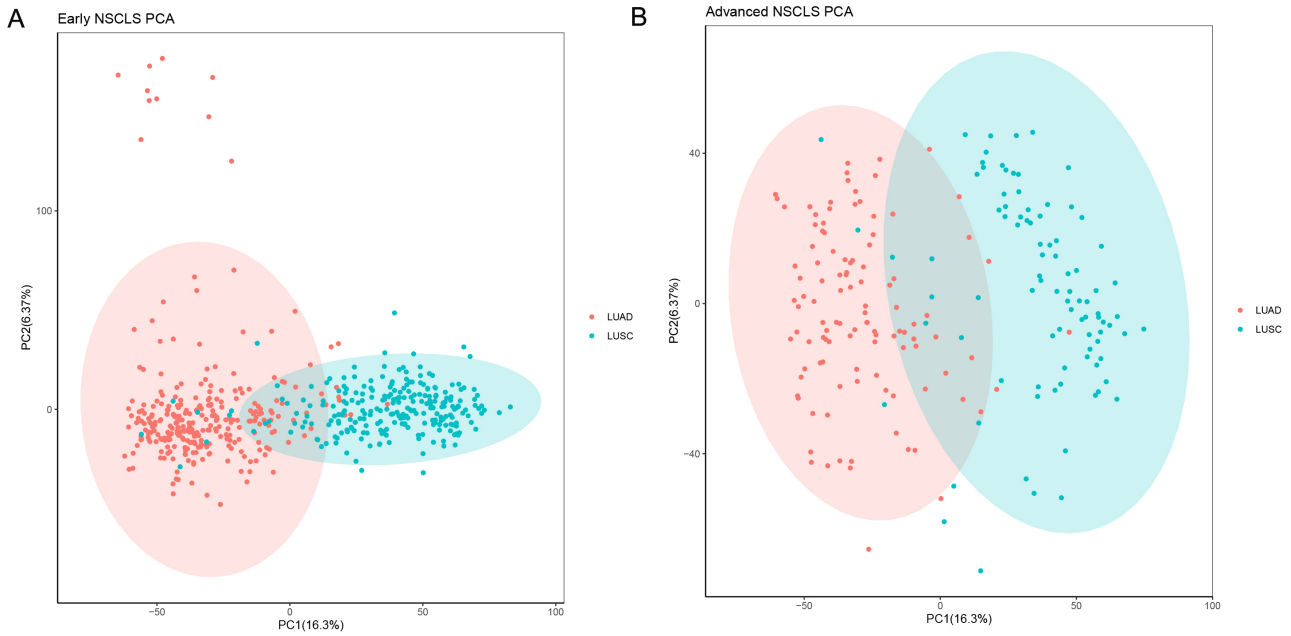

Fig. 1. Visualization of gene expression profile data quality control. (A,B) A PCA plot of the data showing no batch effect in the TCGA NSCLC dataset. Red nodes represent the LUAD cluster, while blue nodes represent the LUSC cluster. PCA, principal component analysis; NSCLC, non-small cell lung carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; TCGA, The Cancer Genome Atlas.

pedia of Genes and Genomes (KEGG) (https://www.kegg.jp/). Metascape TRRUST analysis [18] was used to identify enriched transcription factors among the DEGs. The statistical significance of overrepresented pathways was assessed using an adjusted p-value < 0.05.

Functional Enrichment Based on Functional Class Scoring

The utilization of gene expression profile data was significantly improved when functional class score (FCS), a second-generation functional enrichment method, was compared to ORA [17]. Gene set enrichment analysis (GSEA) was conducted utilizing MSigDB H: hallmark gene sets (50 available gene sets, V7.4) via the "plotGseaTable" function in the "fgsea" R package (http://www.rproject.org/) with parameters set as minSize = 5, maxSize = 1000, and nperm = 10000 [19]. The gene-pathway data file was acquired (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). Gene sets were considered significantly enriched if the false discovery rate (FDR) threshold was <0.05.

Analysis of Immune Cell Infiltration

Single-sample gene set enrichment analysis (ssGSEA) (http://software.broadinstitute.org/gsea/msigdb/index.jsp) was applied to determine immune infiltration, and the normalized ssGSEA data were compared with gene sets using the "GSVA" (R package, http://www.rproject.org/) function. The infiltration levels of 22 immune cell types (B cells naive, B cells memory, plasma cells, T cells CD8, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory, T cells gamma delta, NK cells resting, NK cells activated, monocytes, macrophages M0, macrophages M1, macrophages M2, dendritic cells resting, dendritic cells activated, mast cells resting, mast cells activated, eosinophils, and neutrophils) were estimated using ssGSEA analysis. A boxplot was generated using the "ggplot2" R package (http://www.rproject.org/) to exhibit the differences in the abundance of these 22 infiltrating immune cells between early and advanced stages of NSCLC.

Identification of Co-Expression Immune Cell Infiltration-Related Modules Using WGCNA Algorithm

The WGCNA strategy [12] was employed to detect gene modules associated with immune cell abundance in the early and late stages of NSCLC. Initially, we curated the dataset consisting of DEGs and deleted the outlier samples and genes with missing data, as well as those with zero variance. Subsequently, we applied the "goodSamplesGenes" function from the WGCNA R package (http://www.rproject.org/) for network construction. The optimal soft threshold power for the gene expression profile was determined using the PickSoftThreshold function and sft$powerEstimate parameters of the WGCNA package. According to the power-value, a GeneTree was constructed. Dynamic modules were determined utilizing a minimum size of 30 genes, and highly similar modules were

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

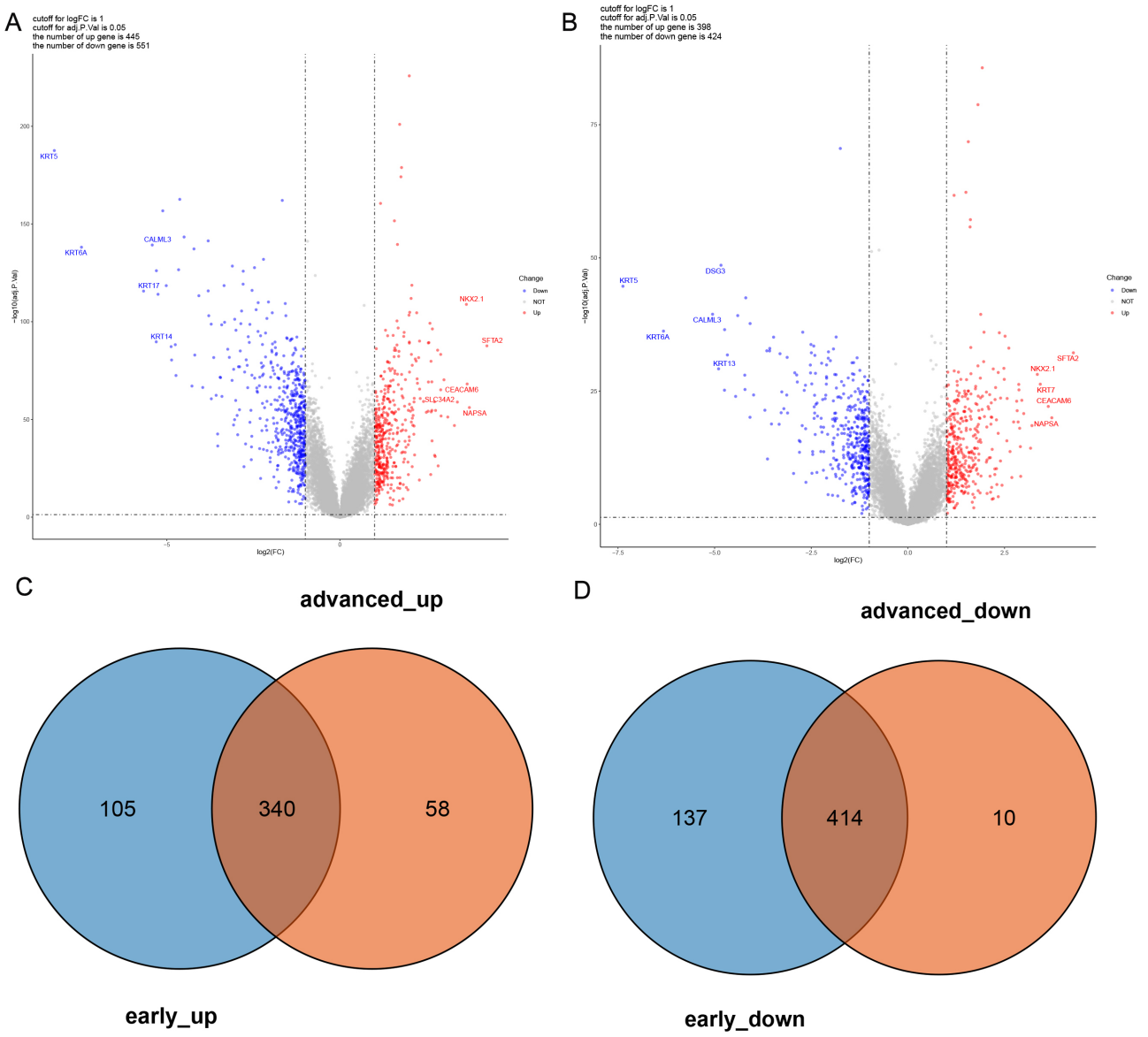

Fig. 2. Visualization of the results of gene differential expression analysis. (A,B) Volcano plot of all differentially expressed genes between LUAD and LUSC in early and advanced NSCLC. Differential gene expression was performed with "limma" and "linear" models. The false discovery rate (FDR)-adjusted *p*-value was used for the plot. (C,D) Venn plot for the differentially expressed genes between LUAD and LUSC in early and advanced NSCLC.

combined using a Diss Threshold of 0.25. The blockwiseModules function was used to construct the network with the following parameters: networkType = "signed", TOMType = "signed", corType = "bicor", minModuleSize = 30, mergeCutHeight = 0.25, and minKMEtoStay = 0.8. Subsequently, we plotted the correlations between modules and traits, including immune infiltration score and co-expressed gene modules, using R software (http://www.rproject.org/). The heatmap of module-trait correlations was generated utilizing the "labeledHeatmap" function. Modules exhibiting the strongest connections with immune cell infiltration were identified, and an intersect evaluation (consensus genes) was performed for early and advanced NSCLC.

Drug Sensitivity and Pathway Activity Linked to Consensus Genes

An examination of drug resistance associated with consensus genes was conducted using the "drug sensitivity analysis" and "pathway activity" modules of "Gene Set Cancer Analysis Literate (GSCALite)" (http://bioinfo.life.hust.edu.cn/web/GSCALite/) [20].

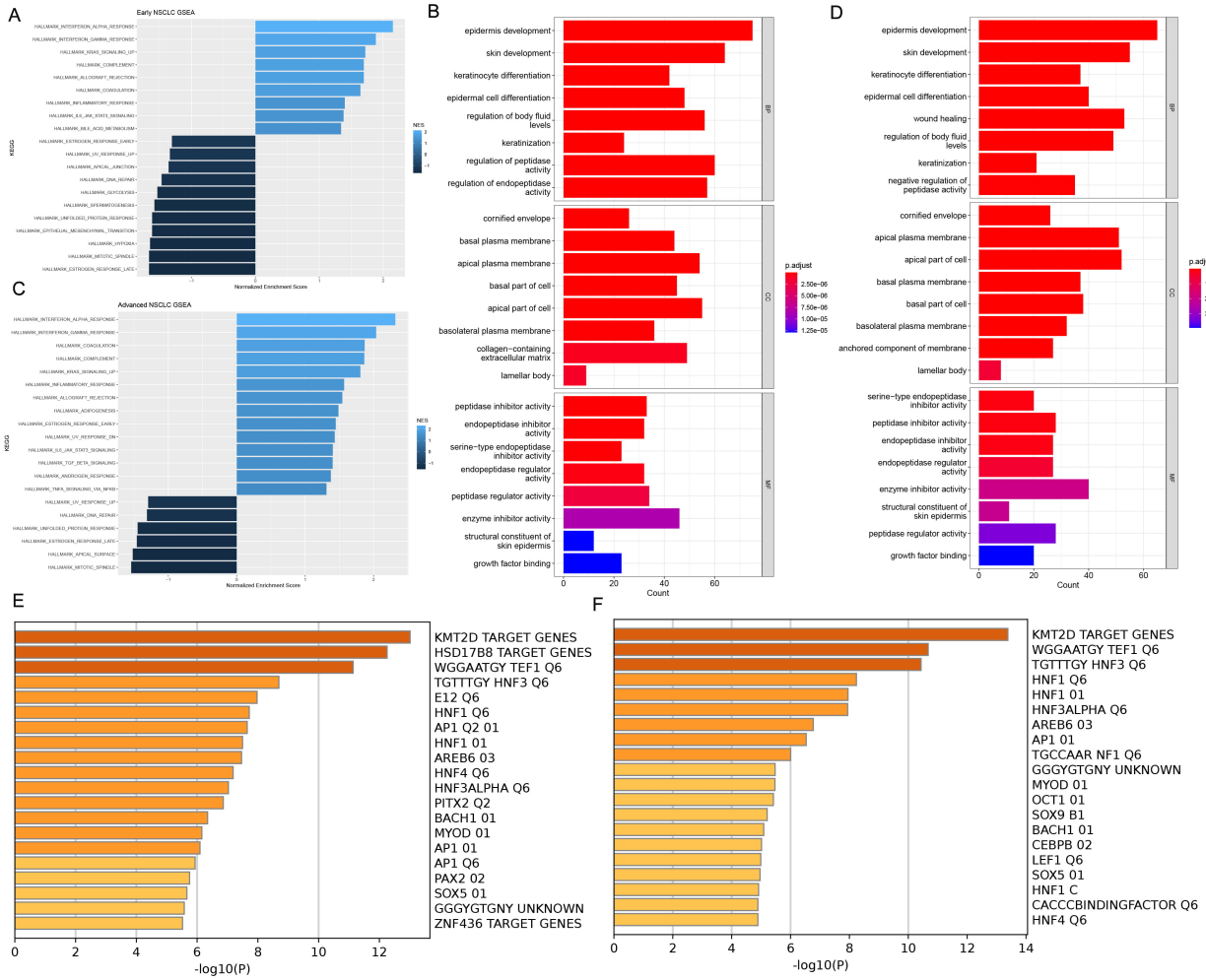

Fig. 3. Differentially expressed genes (DEGs) enriched for IFN signaling. (A,C) Functional enrichment for GSEA of differentially expressed genes in initial and progressive stages of NSCLC. (B,D) Functional enrichment for GO and pathway of differentially expressed genes in initial and progressive stages of NSCLC. (E,F) Functional enrichment of the transcription factors of differentially expressed genes in early and advanced NSCLC. GSEA, gene set enrichment analysis; GO, gene ontology.

Differential Gene Expression Analyzed According to TP53 Mutation Status

We used the "coBarplot" function from the "maftools" R package (http://www.rproject.org/) [21] to compare and extract the significantly mutated genes in LUAD compared to LUSC at both the initial and progressive stages of NSCLC. Genes were screened based on their significant expression differences between groups ($|\log_2 FC| > 1$, adjusted $p < 0.05$).

Survival Analysis for Genes Based on the Kaplan-Meier Platform

Survival analysis was performed to assess the correlation of several consensus genes in individuals with lung cancer utilizing an online Kaplan-Meier plotter [22] (http://kmplot.com/analysis/index.php?p=service&cancer=lung). Additionally, individuals with lung cancer used overall survival (OS) as a measure of their progress.

Statistical Analysis

Quantitative assessments were conducted using online bioinformatics tools, R software (http://www.rproject.org/), and Microsoft Excel Software (version 2018, Microsoft, Redmond, WA, USA). The "VennDiagram" R program was used to create a customizable Venn diagram [23]. The association between modules and clinical characteristics was analyzed using the Pearson correlation coefficient. All *p*-values were derived from two-sided tests, with statistical significance set at $p < 0.05$.

Results

Different Tissue Subtypes of Early and Advanced NSCLC Have Different Gene Expression Profiles

We first performed differential gene expression analyses comparing early and advanced NSCLC (LUAD *vs* LUSC samples) using the "limma" R package. Following that, we reduced the dataset's dimensionality through prin-

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

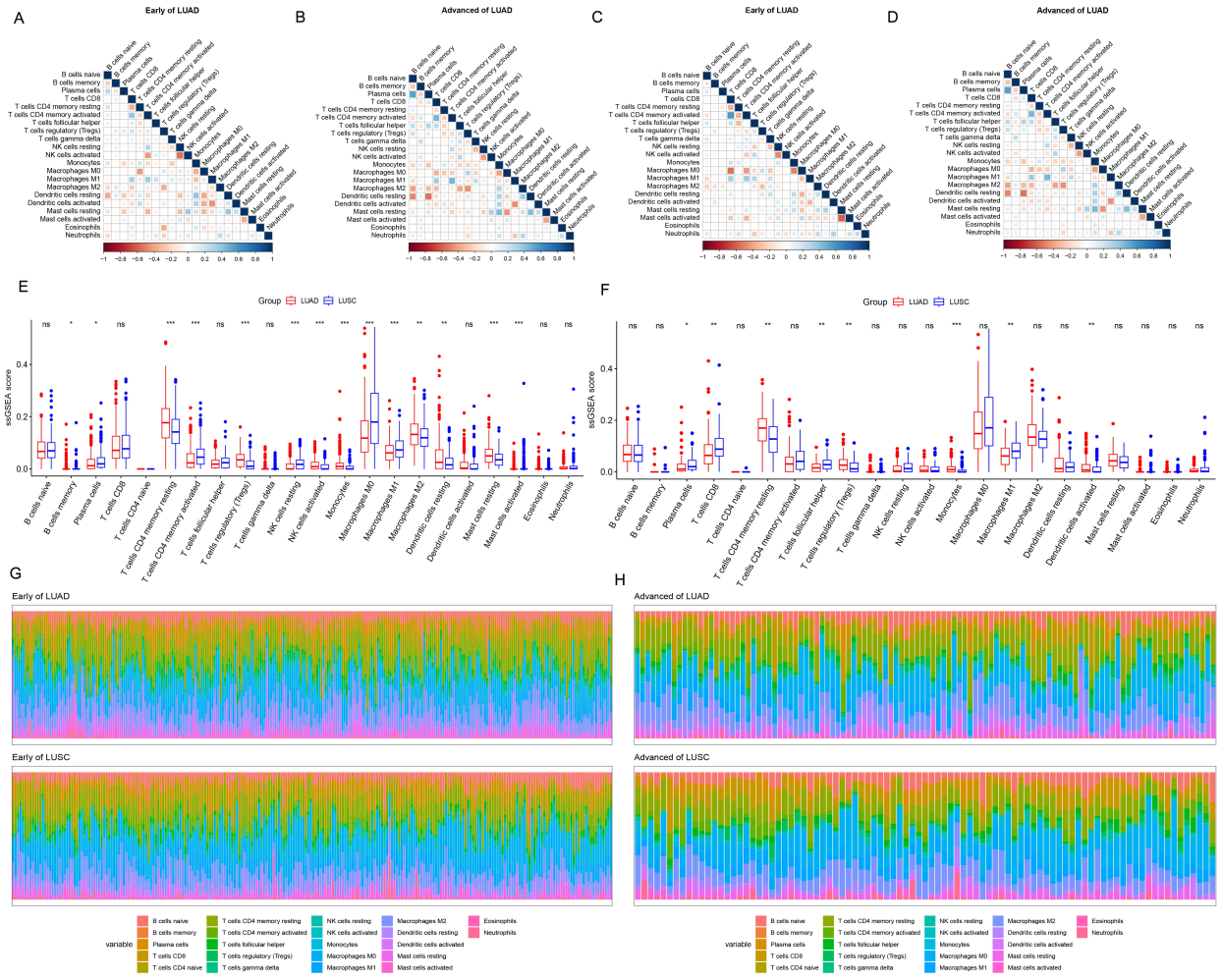

Fig. 4. Immune cell infiltration in NSCLC. (A–D) The correlation between early and advanced NSCLC (LUAD and LUSC) and 22 immune cells. (E,F) Differences in the abundance of immune infiltration of 22 immune cells between early and advanced LUAD and LUSC. (G,H) Percentages of 22 immune cell infiltration abundance in early and advanced LUAD *vs* LUSC. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, statistically significant; ns, no statistical difference.

cipal component analysis (PCA). The two sample groups were normalized for further differential expression analysis (Fig. 1A,B, **Supplementary Fig. 1A,B**). The volcano plot was employed to visualize the DEGs. In the early NSCLC category, the top 5 upregulated and downregulated genes were *NKX2.1, SFTA2, CEACAM6, SLC34A2, NAPSA*, and *KRT5, CALML3, KRT6A, KRT17, KRT14*, respectively (Fig. 2A). For advanced NSCLC, *SFTA2, NKX2.1, KRT7, CEACAM6,* and *NAPSA* were the top 5 upregulated genes, while *DSG3, KRA5, CALML3, KRT6A*, and *KRT13* were the top 5 downregulated genes (Fig. 2B). The comparison of downregulated genes within the DEGs in both early and advanced NSCLC groups revealed a significant overlap, indicating that the differences in expression profiles between these groups predominantly stemmed from upregulated genes. Furthermore, the Venn diagram (Fig. 2C,D) showed that there are 340 commonly upregulated genes and 414 commonly downregulated genes shared between early and advanced NSCLC.

Differentially Expressed Genes are Significantly Enriched in Immune-Related Pathways

The early NSCLC group showed that upregulated genes were significantly enriched in "interferon_alpha_response", "interferon_gamma_response", "kras_signaling_up", and "complement", according to GSEA and GO/Pathway analysis (Fig. 3A,B). Analysis of the advanced NSCLC group indicated that differentially expressed genes were linked to "interferon_alpha_response", "interferon_gamma_response", "coagulation", and "complement" (Fig. 3C,D). In addition, Metascape TRRUST analysis of all statistically enriched transcription factors in the early NSCLC group identified *KMT2D*, *HSD17B8*, and *WGGAATGY TEF1 Q6* as key regulators of gene expression (Fig. 3E). *KMT2D, WGGAATGY TEF1 Q6*, and *TGTTTGY HNF3 Q6* transcription factors have the ability to regulate all significantly enriched transcription factors in advanced NSCLC, according to the Metascape TRRUST analysis (Fig. 3F). The finding demonstrated

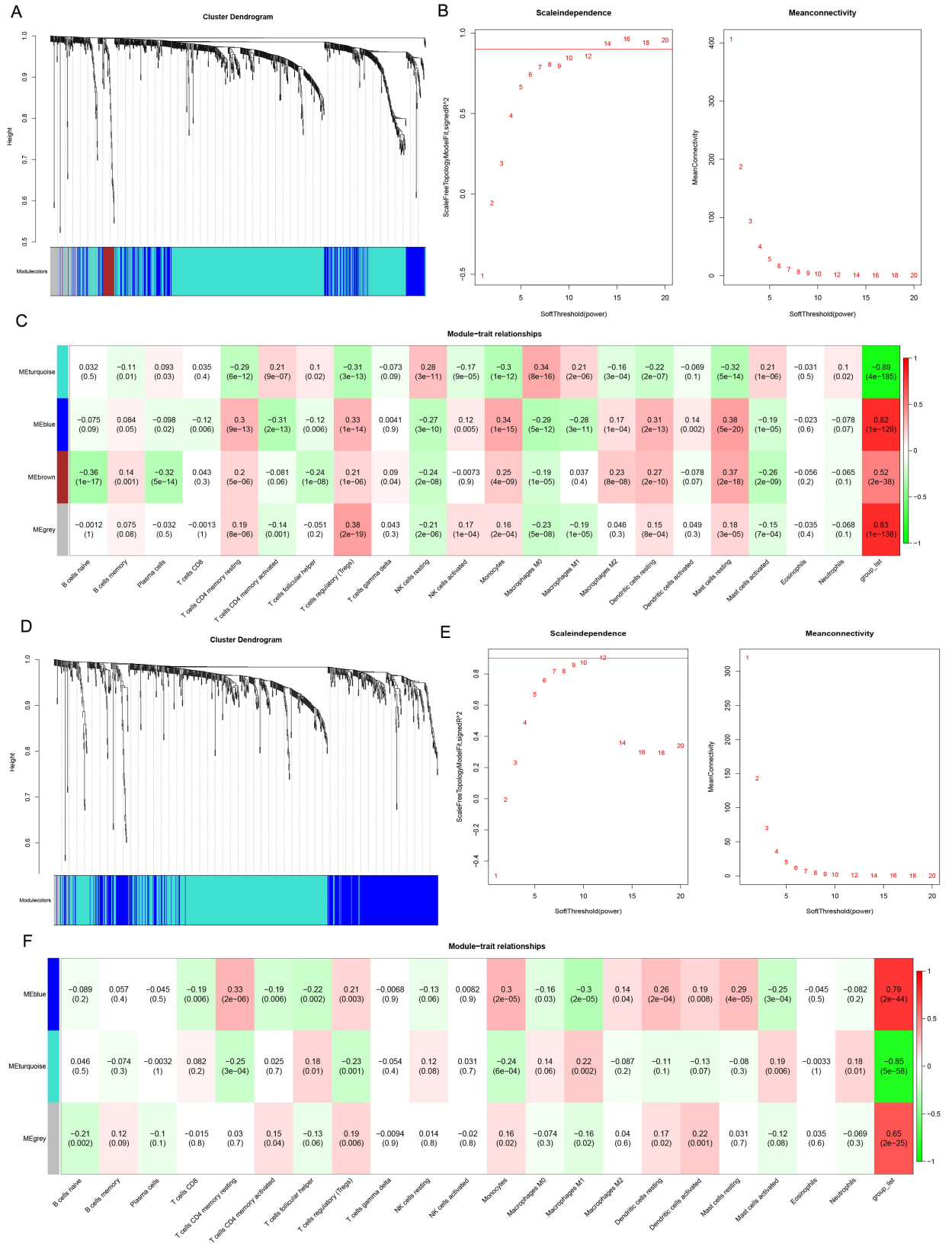

Fig. 5. Development of the weight gene co-expression network. (A,D) Gene dendrogram, with modules represented by different colors. (B,E) Analyses of network topology for various soft-thresholding powers. (C,F) The weighted gene co-expression network analysis identified groups of co-regulated genes (modules). Each square in the figures represents the connection between a module and the corresponding sample, and the *p*-value indicates the significance of the correlation. The colors of the squares indicate different correlation types: positive correlation (red), negative correlation (green), and no correlation (white).

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

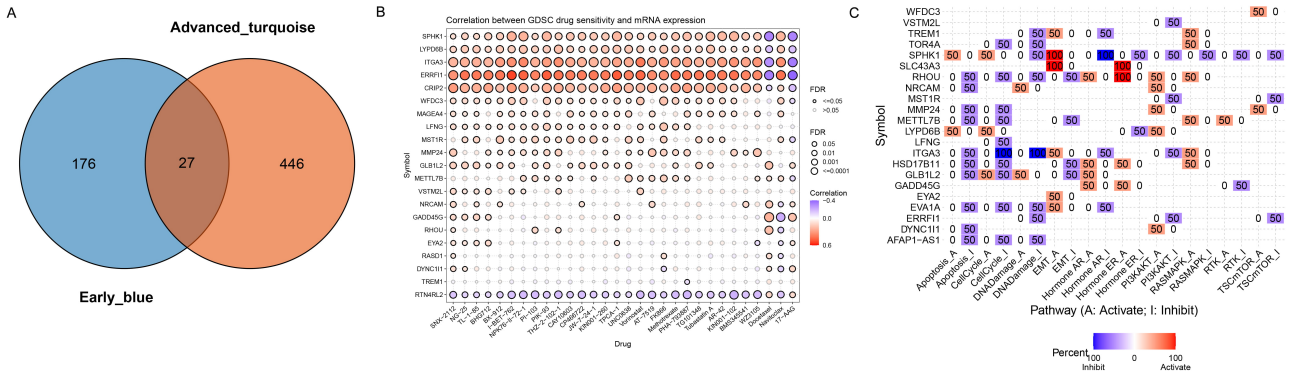

Fig. 6. The association analysis between consensus genes, drug sensitivity, and pathway activity. (A) Consensus genes were identified by Venn plot. (B) The correlation between consensus genes and drug sensitivity was assessed using the Gene Set Cancer Analysis Literate (GSCALite) platform. Red refers to positive correlation, which means the higher the gene expression, the more sensitive to the drug, while blue represents the opposite. (C) Gene-pathway interaction network of consensus genes in NSCLC using the GSCALite platform.

that differentially upregulated genes exhibit considerable enrichment in immune-related biological processes and are controlled by immune-related transcription factors.

Analysis of Immune Cells Infiltration in NSCLC

We investigated the role of immune infiltration in both the early and advanced stages of NSCLC. Infiltration involves the migration and accumulation of immune cells within the tumor microenvironment. Our study utilized ssGSEA analysis to assess immune infiltration in the early and advanced stages of NSCLC. Fig. 4A–D illustrate the connection between early and advanced NSCLC (LUAD and LUSC) and 22 types of immune cells. Additionally, Fig. 4E,F present box plots highlighting the differences in immune cell abundance between early and advanced stages of LUAD and LUSC. Dendritic cells showed the highest levels of infiltration among all immune cell types in both early and advanced stages of NSCLC. Furthermore, the levels of infiltration of dendritic cells, memory-activated CD4 T cells, M2 macrophages, and neutrophils were significantly higher in advanced NSCLC compared to early-stage disease, indicating an increased immune response in later stages. In contrast, the abundance of CD8 T cells was higher in early-stage NSCLC. In Fig. 4G,H, the bars represent the percentages of 22 immune cell infiltration abundance in early and advanced LUAD *vs* LUSC. While there were similarities in the overall patterns between LUAD and LUSC, some differences could be observed, such as a slightly higher neutrophil infiltration in LUSC compared to LUAD in both early and advanced stages.

Identification of Immune Cell-Associated Gene Modules

Immune cell scores were examined using the ssGSEA algorithm. The study utilized WGCNA to establish a model of immune-related genes. Fig. 5A,B show the module-trait cluster dendrogram and soft threshold for the early NSCLC group. In Fig. 5C, the module-trait cluster correlation heatmap revealed that the blue module exhibited the strongest link to the overall immune cell score. Fig. 5D,E present the module-trait cluster dendrogram and soft threshold advanced NSCLC. The module-trait correlation heatmap in Fig. 5F indicated that the turquoise module displayed the strongest correlation with the immune cell score. Finally, 27 "consensus genes" were identified and selected for further analysis by intersecting genes in the blue module of early NSCLC and the turquoise module of advanced NSCLC (Fig. 6A).

The Association between Consensus Genes, Drug Sensitivity, and Pathway Activity

Fig. 6A,B show a strong correlation between ERRFI1 expression and the majority of anti-cancer drugs among the 27 consensus genes. Conversely, *RTN4RL2* yielded contrasting results. In addition, an analysis of gene-pathway activity among the 27 consensus genes indicated that certain genes may contribute to the progression of both early and advanced NSCLC through the RTK, PI3K/AKT, RAS/MAPK and TSCmTOR signaling pathways (Fig. 6C).

Differences in Gene Expression Depending on the TP53 Mutation Status

Consensus genes (*ARHGAP40, MST1R, CRIP2, GADD45G, ITGA3, NRCAM, RASD1, RTN4RL2, TOR4A,* melanoma-associated antigen A 4 (*MAGEA4*), dynein cytoplasmic 1 intermediate chain 1 (*DYNC1I1*) and *LYPD6B*) showed notable variations in gene expression in patients

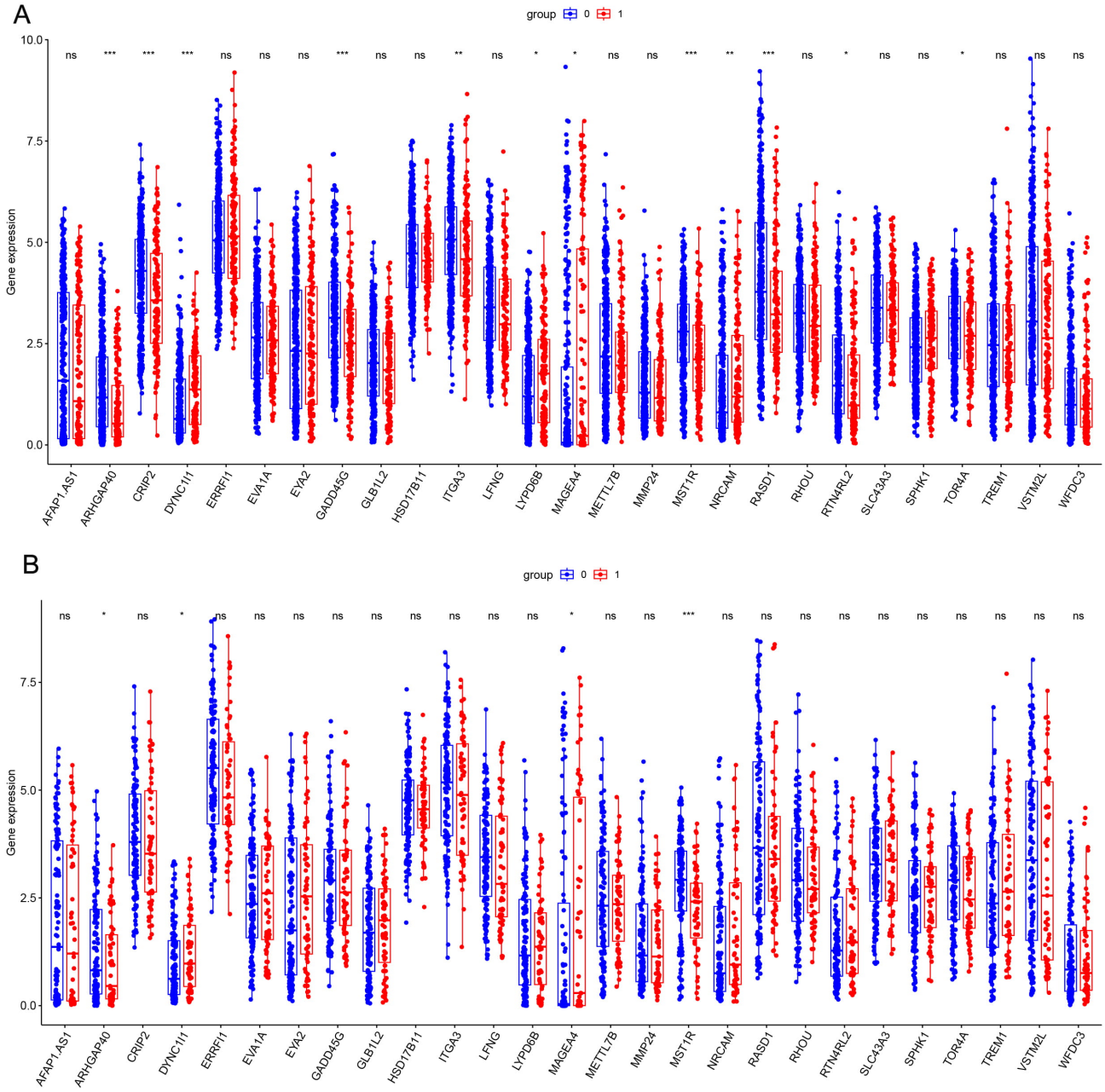

Fig. 7. Expression of consensus genes in different TP53 mutation statuses. (A) Expression of consensus genes in different TP53 mutation statuses in early NSCLC. (B) Expression of consensus genes in different TP53 mutation statuses in advanced NSCLC. *$p <$ 0.05, **$p < 0.01$, ***$p < 0.001$, 0: TP53 wt, 1: TP53 mut, statistically significant; ns, no statistical difference.

with different TP53 mutant statuses (Fig. 7A,B). The identified consensus genes play diverse functional roles related to cancer progression and metastasis. For example, genes such as *ARHGAP40, DYNC1I1*, and *ITGA3* are involved in regulating cell movement, migration, and adhesion to the extracellular matrix [24–26]. Other genes like *MAGEA4* and *NRCAM* may participate in immune response, neuronal migration, axon formation, and nervous system development. Additionally, *GADD45G* and *MST1R* are pivotal in regulating cell proliferation, cell cycle progression, and DNA repair. Consensus genes such as *RASD1* and *CRIP2* are involved in modulating cell signaling pathways, cytoskeleton dynamics and organelle arrangement. Furthermore, *MST1R* and *LYPD6B* may facilitate tumor growth, metastasis, and tissue remodeling. Notably, *DYNC1I1* and *MAGEA4* were selected for subsequent survival analysis due to their increased expression in the mutant group in early and advanced stages of NSCLC.

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

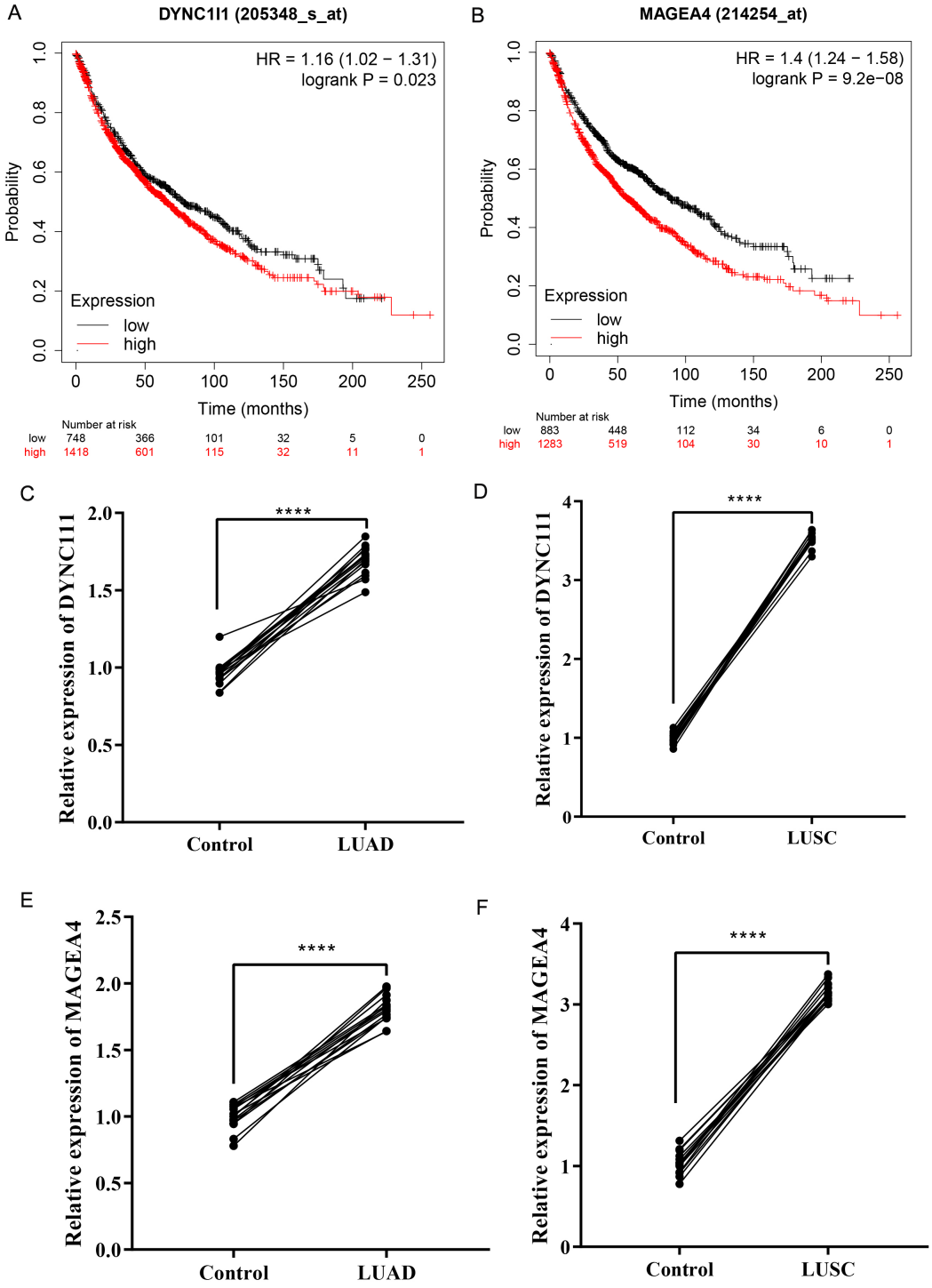

Fig. 8. Melanoma-associated antigen A 4 (*MAGEA4*) and dynein cytoplasmic 1 intermediate chain 1 (*DYNC1I1*) mRNA expressions are positively correlated with a favorable prognosis of NSCLC. (A,B) Kaplan-Meier survival curves of *MAGEA4* and *DYNC1I1* mRNA expression. (C–F) The quantitative polymerase chain reaction (qPCR) for the expression of *MAGEA4* and *DYNC1I1* in NSCLC patients. ****$p < 0.0001$.

Prognostic Potential of Consensus Genes with Different Correlation Patterns in NSCLC

The Kaplan-Meier survival analysis results revealed a significant correlation between elevated expression levels of *MAGEA4* and *DYNC1I1* and poor OS in lung cancer patients (Fig. 8A,B). The quantitative polymerase chain reaction (qPCR) results indicated a substantial upregulation of *MAGEA4* and *DYNC1I1* in NSCLC patient tumor tissues compared to normal controls. This suggests that *MAGEA4* and *DYNC1I1* are highly expressed and activated during the

onset and progression of NSCLC. These findings demonstrate their strong correlation with the malignant progression of NSCLC, making them pivotal biomarkers for assessing prognosis and disease progression (Fig. 8C–F).

Discussion

NSCLC is the leading cause of cancer-related deaths worldwide, primarily attributed to the accumulation of multiple genetic alterations. However, the precise mechanisms governing its tumorigenesis and progression remain poorly understood. The observed number of DEGs in this study aligns with the findings of Xiao J *et al*. [27]. Additionally, distinguishing between early and advanced NSCLC based on gene expression profile data was challenging, suggesting the involvement of distinct mechanisms.

Furthermore, GO and pathway analysis revealed that DEGs were associated with various immune cell activations and the human complement system, which are critical in the tumorigenesis and growth of early and advanced NSCLC. Additionally, GSEA indicated the significant involvement of DEGs in the "interferon_alpha_response" and "interferon_gamma_response" pathways. Remarkably, *SLC34A2* exhibited distinct expression patterns in early-stage NSCLC, while this difference was absent in advanced stage. Additionally, our study revealed a higher abundance of dendritic cells (DC), key antigen-presenting cells, compared to other immune cell types. These findings collectively suggest that alterations in the immune system play an important role in the development of both early and advanced NSCLC.

Next, we identified two consensus gene modules associated with tumor-infiltrating immune cells in early and advanced NSCLC through WGCNA. A moderately strong positive correlation was observed between ERRFI1 expression and the majority of anti-cancer drugs using GSCALite to measure drug sensitivity. Conversely, another gene, *RTN4RL2*, yielded contrasting results. Additionally, these consensus genes may play a role in advancing early and advanced NSCLC through the RTK, PI3K/AKT, RAS/MAPK, and TSCmTOR signaling pathways, as indicated by GSCALite's analysis of gene-pathway interactions.

In a mutational spectrum analysis, it was found that the most common sequence mutation in both early and advanced NSCLC is the TP53 mutation. In different TP53 mutation statuses, the gene expression levels of *ARHGAP40, DYNC1I1, MAGEA4, MST1R, CRIP2, GADD45G, ITGA3, NRCAM, RASD1, RTN4RL2, TOR4A*, and *LYPD6B* showed significant differences based on gene expression analyses. The elevated expression of *MAGEA4* and *DYNC1I1* mRNA was linked to a poor prognosis for NSCLC patients, as revealed by database analysis using KMplotter.

However, our study does have certain drawbacks. First, a larger sample size would produce more accurate results compared to the dataset collected from NSCLC patients in the TCGA. Second, the majority of findings were based on bioinformatics calculations without experimental validation. Therefore, further research using a substantial cohort and experimental methods is necessary to verify the findings of the current investigation.

Conclusions

In summary, our study showed that immune cell infiltration, particularly dendritic cells, displayed significantly high levels of infiltration in both early and advanced NSCLC, indicating their crucial role in tumor development and progression. Our research presents potential tumor immune targets for the treatment of NSCLC.

Abbreviations

TCGA, The Cancer Genome Atlas; NSCLC, non-small cell lung carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; DEGs, differentially expressed genes; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological processes; CC, cellular components; MF, molecular functions; GSEA, gene set enrichment analysis; WGCNA, weighted gene co-expression network analysis; TIMER, tumor immune estimation resource; ssGSEA, single-sample gene set enrichment analysis.

Availability of Data and Materials

All data included in this study can be obtained by contacting the first author if needed.

Author Contributions

WZ designed the research study. HT and ZG performed the research. XM, CL, JL provided help and advice on the experiments and data. WZ, ZG analyzed the data. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

We are thankful to the TCGA Research Network (http://cancergenome.nih.gov/) for providing the data analyzed

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

in our study. This study met the publication guidelines stated by TCGA (https://cancergenome.nih.Gov/publications/publicationguidelines).

Funding

This study is funded by the Innovation Incentive Project of Qiqihar Science and Technology Plan (CSFGG-2022167).

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.23812/j.biol.regul.homeost.agents.20243806.385.

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: a Cancer Journal for Clinicians. 2021; 71: 209–249.

[2] Xiao Y, Cong M, Li J, He D, Wu Q, Tian P, *et al*. Cathepsin C promotes breast cancer lung metastasis by modulating neutrophil infiltration and neutrophil extracellular trap formation. Cancer Cell. 2021; 39: 423–437.e7.

[3] Jemal A, Miller KD, Ma J, Siegel RL, Fedewa SA, Islami F, *et al*. Higher Lung Cancer Incidence in Young Women Than Young Men in the United States. The New England Journal of Medicine. 2018; 378: 1999–2009.

[4] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, *et al*. Cancer statistics in China, 2015. CA: a Cancer Journal for Clinicians. 2016; 66: 115–132.

[5] Jin T, Nguyen ND, Talos F, Wang D. ECMarker: interpretable machine learning model identifies gene expression biomarkers predicting clinical outcomes and reveals molecular mechanisms of human disease in early stages. Bioinformatics (Oxford, England). 2021; 37: 1115–1124.

[6] Guerrera F, Errico L, Evangelista A, Filosso PL, Ruffini E, Lisi E, *et al*. Exploring Stage I non-small-cell lung cancer: development of a prognostic model predicting 5-year survival after surgical resection†. European Journal of Cardio-thoracic Surgery: Official Journal of the European Association for Cardio-thoracic Surgery. 2015; 47: 1037–1043.

[7] Onoi K, Chihara Y, Uchino J, Shimamoto T, Morimoto Y, Iwasaku M, *et al*. Immune Checkpoint Inhibitors for Lung Cancer Treatment: A Review. Journal of Clinical Medicine. 2020; 9: 1362.

[8] Wang Y, Wenzl K, Manske MK, Asmann YW, Sarangi V, Greipp PT, *et al*. Amplification of 9p24.1 in diffuse large B-cell lymphoma identifies a unique subset of cases that resemble primary mediastinal large B-cell lymphoma. Blood Cancer Journal. 2019; 9: 73.

[9] Osipov A, Lim SJ, Popovic A, Azad NS, Laheru DA, Zheng L, *et al*. Tumor Mutational Burden, Toxicity, and Response of Immune Checkpoint Inhibitors Targeting PD(L)1, CTLA-4, and Combination: A Meta-regression Analysis. Clinical Cancer Research: an Official Journal of the American Association for Cancer Research. 2020; 26: 4842–4851.

[10] Meehan K, Leslie C, Lucas M, Jacques A, Mirzai B, Lim J, *et al*. Characterization of the immune profile of oral tongue squamous cell carcinomas with advancing disease. Cancer Medicine. 2020; 9: 4791–4807.

[11] 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 Research. 2017; 77: e108–e110.

[12] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9: 559.

[13] Bruford EA, Braschi B, Denny P, Jones TEM, Seal RL, Tweedie S. Guidelines for human gene nomenclature. Nature Genetics. 2020; 52: 754–758.

[14] Shi R, Bao X, Unger K, Sun J, Lu S, Manapov F, *et al*. Identification and validation of hypoxia-derived gene signatures to predict clinical outcomes and therapeutic responses in stage I lung adenocarcinoma patients. Theranostics. 2021; 11: 5061–5076.

[15] Blakely CM, Watkins TBK, Wu W, Gini B, Chabon JJ, McCoach CE, *et al*. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. Nature Genetics. 2017; 49: 1693–1704.

[16] Sima M, Vrbova K, Zavodna T, Honkova K, Chvojkova I, Ambroz A, *et al*. The Differential Effect of Carbon Dots on Gene Expression and DNA Methylation of Human Embryonic Lung Fibroblasts as a Function of Surface Charge and Dose. International Journal of Molecular Sciences. 2020; 21: 4763.

[17] Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Computational Biology. 2012; 8: e1002375.

[18] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, *et al*. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications. 2019; 10: 1523.

[19] Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2021. (preprint)

[20] Liu CJ, Hu FF, Xia MX, Han L, Zhang Q, Guo AY. GSCALite: a web server for gene set cancer analysis. Bioinformatics (Oxford, England). 2018; 34: 3771–3772.

[21] Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Research. 2018; 28: 1747–1756.

[22] Győrffy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PloS One. 2013; 8: e82241.

[23] Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011; 12: 35.

[24] Wang J, Liu D, Gu Y, Zhou H, Li H, Shen X, *et al*. Potential prognostic markers and significant lncRNA-mRNA co-expression pairs in laryngeal squamous cell carcinoma. Open Life Sciences. 2021; 16: 544–557.

[25] Gong LB, Wen T, Li Z, Xin X, Che XF, Wang J, *et al*. DYNC1I1 Promotes the Proliferation and Migration of Gastric Cancer by Up-Regulating IL-6 Expression. Frontiers in Oncology. 2019; 9: 491.

[26] Li Y, Li F, Bai X, Li Y, Ni C, Zhao X, *et al*. *ITGA3* Is Associated With Immune Cell Infiltration and Serves as a Favorable Prognostic Biomarker for Breast Cancer. Frontiers in Oncology. 2021; 11: 658547.

[27] Xiao J, Lu X, Chen X, Zou Y, Liu A, Li W, *et al*. Eight potential biomarkers for distinguishing between lung adenocarcinoma and squamous cell carcinoma. Oncotarget. 2017; 8: 71759–71771.