Exploring Hypoxia-Related Genes as Prognostic Indicators in Lung Adenocarcinoma

Bandar Alghamdi[1] and Sonia Rocha[2]*

[1]Institute of Molecular Medicine, University Sains Malaysia, Kelantan. Malaysia.

[2]Department of Biochemistry, Cell and Systems Biology, Institute of Systems Molecular and Integrative Biology, University of Liverpool. Liverpool United Kingdom.

*correspondence to: Sonia Rocha srocha@liverpool.ac.uk

Abstract

Background: Lung adenocarcinoma (LUAD) is a leading cause of cancer-related mortality, with hypoxia contributing to tumor progression and treatment resistance. Identifying hypoxia-related biomarkers could enhance prognosis and therapeutic strategies for LUAD. **Methods:** This study aimed to identify hypoxia-related differentially expressed genes (HRDEGs) in lung adenocarcinoma (LUAD) through differential expression analysis. Functional analysis and protein-protein interaction (PPI) network construction were performed to explore the biological roles and interactions of these genes. Kaplan-Meier survival analysis, univariate Cox regression, and Lasso regression were used to identify key genes associated with survival. Multivariate Cox regression was then conducted to assess independent prognostic factors. **Results:** This analysis revealed 283 upregulated HRDEGs and 322 downregulated HRDEGs in LUAD. Functional enrichment analysis indicated that the upregulated genes were primarily involved in cancer-related and cellular signaling pathways, while downregulated genes were associated with immunity-related pathways. We further identified 201 common upregulated hub genes (including *MMP9, CDH1, HSP90AB1, SOX2, CDKN2A, SPP1, EZH2*) and 224 common downregulated hub genes (such as *IL6, TNF, IL1B, JUN, CCL2, TLR4, FOS, PTGS2*). Kaplan-Meier survival analysis, univariate Cox regression, and Lasso regression led to the identification of 17 key genes (*ADRB2, ALDH2, CAT, CCNE1, MAP3K8, DSG2, EIF6, ABCB1, PIK3R1, RAD51, SFTPD, SOD3, CLEC3B, ADAM12, EXO1, FBLN5,* and *IGF2BP3*) associated with patient survival. Finally, multivariate Cox regression analysis identified *DSG2, EIF6,* and *EXO1* as independent prognostic factors for LUAD, highlighting their potential as biomarkers for prognosis and therapeutic targets in lung cancer. **Conclusion:** In conclusion, *DSG2, EIF6,* and *EXO1* were identified as key hypoxia-related genes in lung adenocarcinoma. These genes were found to be independent

prognostic factors, highlighting their potential as biomarkers for predicting patient survival and guiding future therapeutic approaches.

Keyword: Lung Adenocarcinoma, Hypoxia-Related Differentially Expressed Genes, immunity, independent prognostic factors

Authors summary:

This study investigates the role of hypoxia-related genes in lung adenocarcinoma (LUAD), a major cause of cancer mortality. By analyzing gene expression data, this study identified 283 upregulated and 322 downregulated hypoxia-related differentially expressed genes (HRDEGs). Functional enrichment analysis revealed that the upregulated genes are largely involved in cancer progression and cell signaling, while the downregulated ones are linked to immune responses. A set of key hub genes was identified through protein-protein interaction analysis. Using survival analysis tool, 17 genes significantly associated with patient outcomes in two different patient cohorts. Among these, confirmed three genes—DSG2, EIF6, and EXO1—were confirmed as independent prognostic markers. These findings suggest that these genes may serve as valuable biomarkers for predicting survival and developing targeted therapies in LUAD, offering new avenues for improving patient management in hypoxia-driven tumors.

Introduction

Lung Adenocarcinoma and Hypoxia Lung cancer remains one of the most common and deadly cancers worldwide, with non-small cell lung cancer (NSCLC) being the dominant histological subtype[1]. Lung adenocarcinoma (LUAD) is a subtype of NSCLC and is characterized by distinct molecular and genetic features. Despite significant advances in the understanding of lung cancer biology, the prognosis for patients diagnosed with LUAD remains poor, especially in advanced stages[2]. The overall five-year survival rate for LUAD patients is low, making it essential to identify novel biomarkers for early diagnosis, prognosis, and therapeutic targeting[3]. One of the critical microenvironments influencing cancer progression, metastasis, and resistance to therapy is tumor hypoxia. Hypoxia refers to the state of low oxygen availability within the tumor, often resulting from an imbalance between the rapid growth of tumor cells and the insufficient development of blood vessels to supply oxygen[4]. Hypoxia is a hallmark feature of many solid tumors, including lung adenocarcinoma, and it is associated with aggressive tumor behavior, poor prognosis, and resistance to conventional therapies[5]. The development of therapeutic strategies targeting hypoxic tumor cells remains an area of active research.

The Role of Hypoxia in Tumorigenesis Hypoxia plays a crucial role in driving various aspects of tumor biology, such as tumor angiogenesis, metabolic reprogramming, and epithelial-to-mesenchymal transition (EMT)[6]. It activates a range of adaptive cellular responses through hypoxia-inducible factors (HIFs), primarily HIF-1α and HIF-2α. These transcription factors regulate the expression of genes that mediate critical processes such as angiogenesis (formation of new blood vessels), glycolysis, metastasis, and drug resistance[7]. Tumor cells exposed to hypoxic conditions undergo metabolic shifts that favor anaerobic glycolysis, which allows them to survive and proliferate despite low oxygen availability. The HIF signaling pathway is essential for the

regulation of genes involved in cellular survival, metabolism, and tumor progression[8]. Among the most well-known hypoxia-regulated genes are vascular endothelial growth factor (VEGF), glucose transporter proteins (e.g., GLUT1), and matrix metalloproteinases (e.g., MMPs)[9]. These genes promote tumor growth and metastasis by facilitating the adaptation of tumor cells to low-oxygen conditions. Additionally, hypoxia is implicated in the regulation of immune evasion and resistance to chemotherapies and radiation therapy, further complicating treatment strategies[10]. Therefore, the identification of key hypoxia-related genes and their role in LUAD progression and prognosis is crucial for improving the clinical management of lung adenocarcinoma.

The link between hypoxia and poor prognosis in LUAD has been well established through various studies. Tumors with hypoxic regions have been shown to have a higher potential for metastasis, increased tumor aggressiveness, and reduced sensitivity to treatment[11]. Patients with high expression levels of hypoxia-related genes often exhibit shorter overall survival and progression-free survival compared to those with low expression levels of these genes[12]. Several studies have sought to identify specific hypoxia-related biomarkers that could serve as predictors of patient outcomes in LUAD. However, the heterogeneity of LUAD, coupled with the complexity of hypoxia-induced changes in gene expression, presents challenges in pinpointing specific prognostic genes. Additionally, the tumor microenvironment in LUAD is influenced by various factors, such as genetic mutations, immune responses, and stromal interactions, which may confound the impact of hypoxia on survival.

In this context, a comprehensive analysis of hypoxia-related genes and their potential prognostic value in LUAD is crucial. Identifying genes that are consistently associated with poor prognosis, regardless of other clinical variables, will allow for the development of better prognostic models. These models could guide personalized treatment strategies, such as targeting hypoxic regions of

the tumor with hypoxia-activated prodrugs, or improving the efficacy of therapies through the modulation of the HIF pathway. To achieve this goal, we will utilize a combination of computational and statistical approaches to analyze large-scale genomic datasets.

The analysis will be conducted in a stepwise manner, beginning with a differential expression analysis to identify hypoxia-related genes that are differentially expressed in lung adenocarcinoma. Following this, functional analysis will be performed to explore the biological pathways and roles of these genes. A protein-protein interaction (PPI) network will then be constructed to identify key hub genes involved in hypoxia-related interactions. Kaplan-Meier survival analysis will be conducted to assess the association between the expression of hub genes and overall survival. Univariate Cox regression analysis will evaluate the individual impact of each gene on survival, while Lasso regression will refine the list by selecting genes with non-zero coefficients, eliminating those with no significant impact. Finally, multivariate Cox regression will adjust for potential confounders, providing a more robust estimate of independent prognostic factors. This comprehensive approach aims to identify key hypoxia-related biomarkers that can be used for prognosis in lung adenocarcinoma.

Methods

Data downloading and processing

The analysis utilized two distinct datasets: the TCGA LUAD cohort and GSE18842. The TCGA LUAD cohort comprised 517 tumor samples and 59 control samples, representing lung adenocarcinoma (LUAD) patients and normal controls, respectively. Clinical data from this cohort were incorporated for Kaplan-Meier survival analysis, univariate regression analysis, and multivariate regression analysis to investigate the relationship between gene expression and survival outcomes. For the GSE18842 dataset, which contains 46 tumor samples and 45 control samples. data normalization was performed using a log2 transformation. To account for multiple probes corresponding to individual genes, the data was averaged, resulting in a single expression value for each gene. The GSE18842 dataset is publicly available and can be accessed at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18842.

Differentially expressed genes identifications

In this study, we utilized multiple datasets to identify differentially expressed genes (DEGs) associated with tumorigenesis. The LUAD dataset, consisting of 517 tumor samples, was compared with 59 normal tissue samples from TCGA, while additional data from the GSE18842 dataset, comprising 91 samples (46 tumors and 45 controls), were also incorporated. To identify DEGs between tumor and normal samples, we employed the LIMMA package, applying a stringent cutoff of $|\log2FC| > 0.585$ and an adjusted p-value < 0.05. Following DEG identification, we specifically targeted hypoxia-related differentially expressed genes (HRDEGs), which were selected through a Venn diagram approach. This method allowed us to cross-reference the LUAD DEGs and further refine the set of genes related to hypoxia, thus providing a focused analysis of the genes potentially involved in the tumor microenvironment's hypoxic conditions.

Curation of hypoxia related genes and identification of hypoxia related differentially expressed genes (HRDEGs)

To curate a comprehensive list of hypoxia-related genes, we integrated data from several established resources. First, we retrieved hypoxia-related genes from Gene Set Enrichment Analysis (GSEA) (https://www.gsea-msigdb.org/gsea/index.jsp), focusing on the following gene sets: GOBP HYPOXIA INDUCIBLE FACTOR 1ALPHA SIGNALING PATHWAY, GOBP INTRINSIC APOPTOTIC SIGNALING PATHWAY IN RESPONSE TO HYPOXIA, GOBP NEGATIVE REGULATION OF CELLULAR RESPONSE TO HYPOXIA, GOBP NEGATIVE REGULATION OF HYPOXIA INDUCED INTRINSIC APOPTOTIC SIGNALING PATHWAY,GOBP REGULATION OF CELLULAR RESPONSE TO HYPOXIA, HALLMARK HYPOXIA, and REACTOME REGULATION OF GENE EXPRESSION BY HYPOXIA INDUCIBLE FACTOR. We then performed a search on OMIM (https://www.ncbi.nlm.nih.gov/omim/?term=HYPOXIA) using the term "HYPOXIA" to identify genes associated with hypoxia-related genetic disorders. Additionally, we retrieved genes from the hsa04066: HIF-1 signaling pathway in the KEGG database (https://www.genome.jp/kegg/). Finally, we obtained genes from GeneCards (https://www.genecards.org/), selecting those with a score greater than or equal to 1. After compiling the data from these sources, we performed deduplication to ensure that only unique genes were included. This resulted in a final list of 3292 unique genes, which are tabulated in **Supplementary Table S1**.

Functional enrichment analysis of

We used DAVID v6.8 (https://david.ncifcrf.gov/) [13] to perform functional annotation of the identified genes. This platform allowed us to categorize the genes into various biological contexts,

such as significantly enriched KEGG pathways, biological processes, cellular components, and molecular functions. These functional analyses provided valuable insights into the roles and interactions of the genes in key cellular mechanisms. To assess the statistical significance of the enrichment, we applied a false discovery rate (FDR) threshold of < 0.05, calculated using the Benjamini–Hochberg method [14], ensuring that only robust and biologically relevant findings were included.

Exploration of hub HRDEGs

We constructed Protein-Protein Interaction (PPI) networks for the differentially expressed genes (DEGs) using STRING v11.0 (https://string-db.org/) [15]). Both upregulated and downregulated genes were entered into the STRING search module to generate the interaction networks. To identify key hub genes within these networks, we utilized the cytoHubba plugin in Cytoscape (https://cytoscape.org/) [16], which ranks genes based on their interaction degree. A medium interaction score threshold of ≥ 0.40 was applied to build the PPI network, and genes with an interaction degree of ≥ 10 were selected for further analysis. This approach enabled us to focus on the most significant hub genes, highlighting those with strong interactions relevant to lung cancer.

Exploration of survival associated hub HRDEGs

We retrieved clinical data from TCGA LUAD to perform an analysis of overall survival (OS) in relation to gene expression. Patients were categorized into two groups based on their gene expression levels, either above or below the median, and survival rates between these groups were compared. To further explore survival differences, patients were also stratified into risk groups based on their gene expression profiles. The log-rank test and Kaplan-Meier survival curves were utilized to evaluate and visualize differences in survival outcomes. For a more detailed survival analysis, we used the R package "survival" [17], performing both univariate and multivariate Cox

regression analyses to assess the independent effects of genes on survival. Additionally, we applied LASSO-Cox regression via the "glmnet" package [18], incorporating 10-fold cross-validation to optimize the selection of penalty parameters and minimize the risk of overfitting. This comprehensive approach allowed us to identify significant survival-associated genes while controlling for potential biases and overfitting, ensuring robust and reliable findings.

Result and discussion

Identification of hypoxia related differentially expressed genes (HRDEGs) in Lung adeno carcinoma

We explored differentially expressed genes (DEGs) in the TCGA LUAD cohort (**Supplementary Table S2**) and identified additional DEGs from the GSE18842 dataset (**Supplementary Table S3**). Volcano plots depicting the results of both analyses are shown in Figure 1A and Figure 1B, respectively. These plots provide a visual representation of the relationship between fold changes (logFC) and adjusted p-values, helping to identify genes with significant differential expression in lung adenocarcinoma (LUAD). Next, we compared the identified DEGs with hypoxia-related genes to determine the hypoxia-related differentially expressed genes (HRDEGs). This comparison revealed a total of 283 upregulated HRDEGs and 322 downregulated HRDEGs in LUAD, which are displayed in Figure 1C-D and detailed in **Supplementary Table S4**. These findings indicate that specific hypoxia-related genes are significantly altered in LUAD, providing potential insights into the molecular mechanisms underlying the disease. The top ten upregulated and downregulated HRDEGs are listed in Table 1. Among the top upregulated HRDEGs are *CYP24A1, CA9, MMP13, SPP1, COL10A1, MMP1, EEF1A2, GREM1, PITX1,* and *MMP12.* These genes are associated with various biological processes such as cell metabolism, matrix remodeling, and tumor progression. On the other hand, the top downregulated HRDEGs include *SFTPC, SLC6A4, AGER, FABP4, SCGB1A1, VEGFD, HBB, SFTPA2, AQP4,* and *CLEC3B*, which are involved in processes like lung surfactant production, cell signaling, and vasculature maintenance.

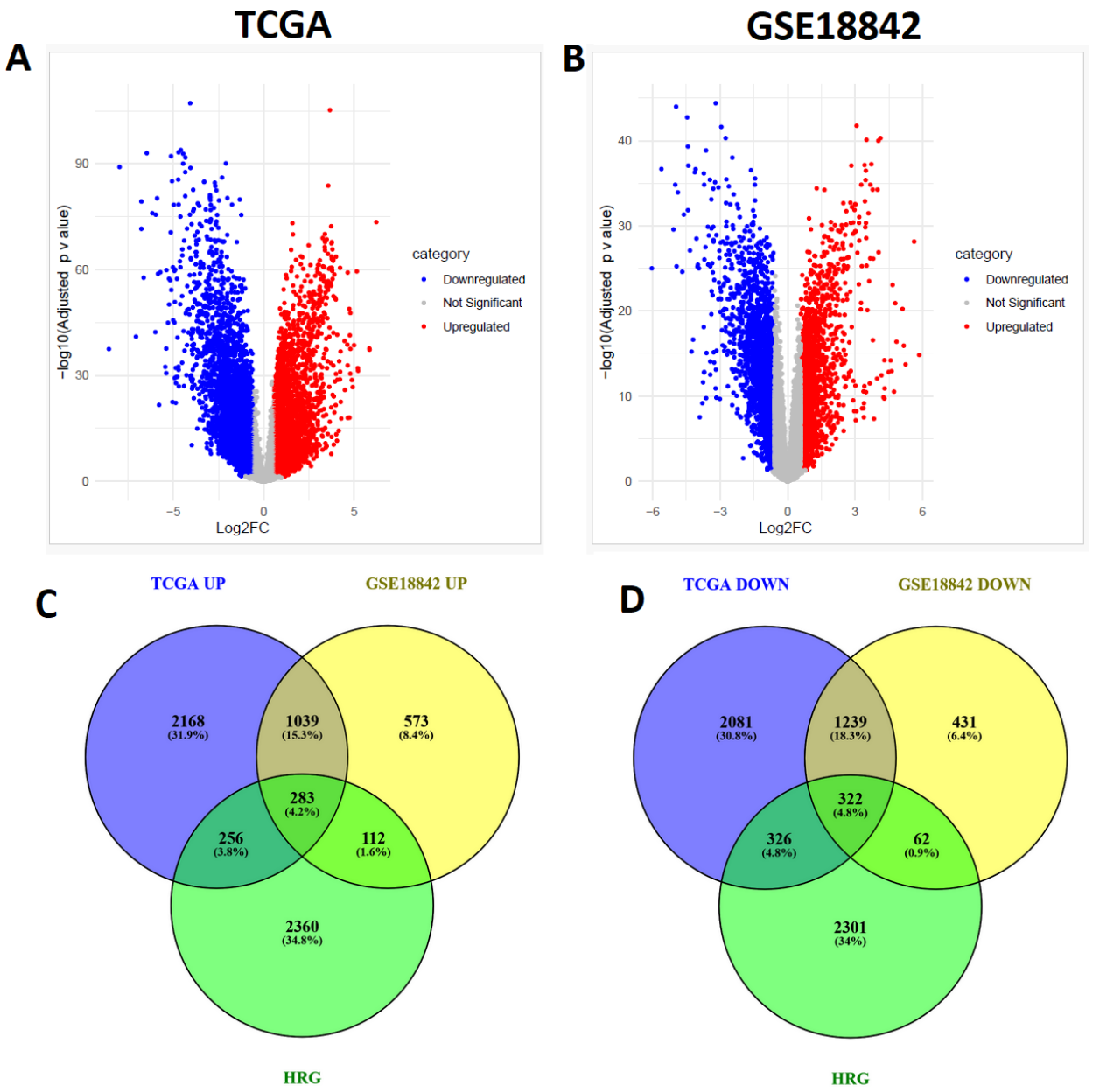


Figure 1. Volcano plots and Venn diagrams depicting differentially expressed genes (DEGs) and hypoxia-related differentially expressed genes (HRDEGs) in lung adenocarcinoma (LUAD). (A) Volcano plot of DEGs in the TCGA LUAD cohort, highlighting genes with significant changes in expression based on log fold change (logFC) and adjusted p-values. (B) Volcano plot of DEGs in the GSE18842 dataset, illustrating differential gene expression in LUAD. (C) Venn diagram showing the overlap of upregulated hypoxia-related differentially expressed genes (HRDEGs) in LUAD. (D) Venn diagram illustrating the overlap of downregulated hypoxia-related differentially expressed genes (HRDEGs) in LUAD.

Table 1. The top ten upregulated and downregulated HRDEGs in TCGA and GSE18842

Regulatory status	Gene ID	Gene Symbol	TCGA		GSE18842		Name of the genes
			logFC	Adjusted P Value	logFC	Adjusted P Value	
Upregulated	1591	*CYP24A1*	5.211	8.81E-33	1.293	9.12E-06	cytochrome P450 family 24 subfamily A member 1
	768	*CA9*	4.929	2.12E-27	1.749	3.30E-13	carbonic anhydrase 9
	4322	*MMP13*	4.885	2.07E-29	1.825	1.41E-07	matrix metallopeptidase 13
	6696	*SPP1*	4.739	1.41E-49	2.540	2.92E-20	secreted phosphoprotein 1
	1300	*COL10A1*	4.653	6.9E-60	3.547	8.02E-21	collagen type X alpha 1 chain
	4312	*MMP1*	4.651	1.47E-33	5.110	5.85E-21	matrix metallopeptidase 1
	1917	*EEF1A2*	4.297	1.69E-18	0.989	2.14E-05	eukaryotic translation elongation factor 1 alpha 2
	26585	*GREM1*	4.257	1.7E-37	4.666	8.94E-24	"gremlin 1, DAN family BMP antagonist"
	5307	*PITX1*	4.233	9.46E-32	1.629	6.43E-17	paired like homeodomain 1
	4321	*MMP12*	4.085	1.72E-29	5.626	7.05E-29	matrix metallopeptidase 12
Downregulated	6440	*SFTPC*	-8.560	3.4E-38	-6.0378	1.01E-25	surfactant protein C
	6532	*SLC6A4*	-7.971	7.8E-90	-1.9304	1.49E-15	solute carrier family 6 member 4
	177	*AGER*	-6.765	2.62E-72	-4.9568	1.01E-44	advanced glycosylation end-product specific receptor
	2167	*FABP4*	-6.160	9.71E-77	-3.4366	5.77E-26	fatty acid binding protein 4
	7356	*SCGB1A1*	-5.786	2.46E-22	-3.9076	3.06E-08	secretoglobin family 1A member 1
	2277	*VEGFD*	-5.008	2.11E-46	-4.4324	4.79E-40	vascular endothelial growth factor D
	3043	*HBB*	-4.916	1.54E-51	-2.5255	1.07E-25	hemoglobin subunit beta
	729238	*SFTPA2*	-4.863	6.09E-23	-3.7907	6.99E-10	surfactant protein A2
	361	*AQP4*	-4.688	1.29E-27	-3.9426	9.98E-26	aquaporin 4
	7123	*CLEC3B*	-4.241	2.27E-72	-4.4159	8.41E-38	C-type lectin domain family 3 member B

HRDEGs are associated with functional enrichment

Figure 2 illustrates the pathways significantly enriched by the upregulated genes, with a false discovery rate (FDR) < 0.05. These pathways highlight key biological processes and molecular mechanisms potentially associated with tumorigenesis (**Supplementary Table S5**). Notably, the cell cycle (hsa04110), biosynthesis of amino acids (hsa01230), and carbon metabolism (hsa01200)

are central to cellular growth and energy production. The p53 signaling pathway (hsa04115) and cellular senescence (hsa04218) are also implicated, reflecting the regulation of cell death and aging processes. Other enriched pathways, such as glycolysis/gluconeogenesis (hsa00010) and HIF-1 signaling (hsa04066), suggest alterations in metabolic pathways and hypoxic responses often observed in cancer. Additionally, pathways related to cancer progression, such as central carbon metabolism in cancer (hsa05230), pathways in cancer (hsa05200), and microRNAs in cancer (hsa05206), are also prominently featured, highlighting the complex molecular interactions in cancer biology. Other enriched pathways include DNA replication (hsa03030), mismatch repair (hsa03430), and glutathione metabolism (hsa00480), which are critical for genome stability and oxidative stress responses. The statistical analysis of upregulated pathways is tabulated in Supplementary Table S5. These results emphasize the multifaceted nature of the upregulated genes and their involvement in various cellular functions relevant to cancer development.

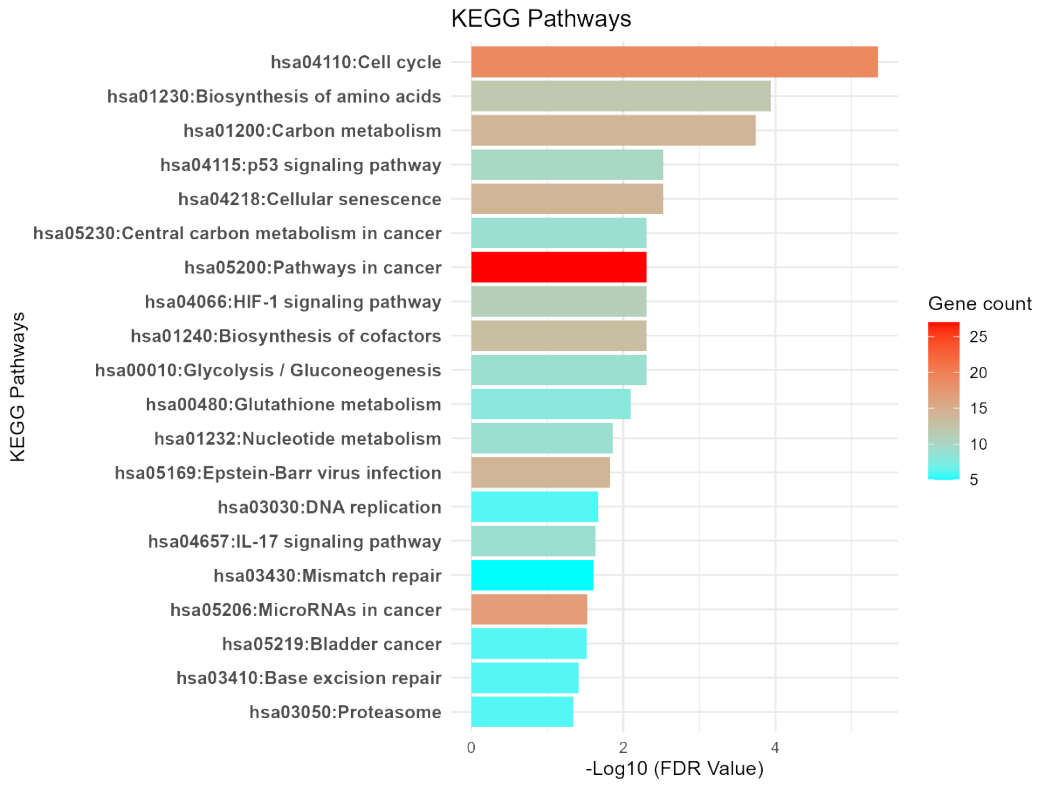


Figure 2. Enriched pathways associated with the upregulated genes (FDR < 0.05). This figure presents a range of biological pathways significantly linked to the upregulated genes, including key processes such as cell cycle regulation, biosynthesis of amino acids, glycolysis, and cancer-related pathways. The results highlight the complex molecular mechanisms underlying tumorigenesis, with a focus on metabolic and cell growth pathways, as well as key signaling pathways like HIF-1.

The top enriched pathway, the Cell Cycle (**Figure 3**), with a highly significant p-value of 4.51E-06. This pathway is critical for regulating cell division and proliferation. The genes identified within this pathway, including *PCNA, MCM7, CDKN2A, PRKDC, HDAC1, CDC6, and others such as CDK4, CHEK1,* and *E2F3*, play pivotal roles in DNA replication, cell cycle progression, and checkpoint control. These genes are integral to maintaining proper cell cycle regulation, and their upregulation suggests a potential disruption in normal cell division processes, which is

commonly associated with tumorigenesis. The enrichment of these genes further underscores the importance of the cell cycle in cancer biology.

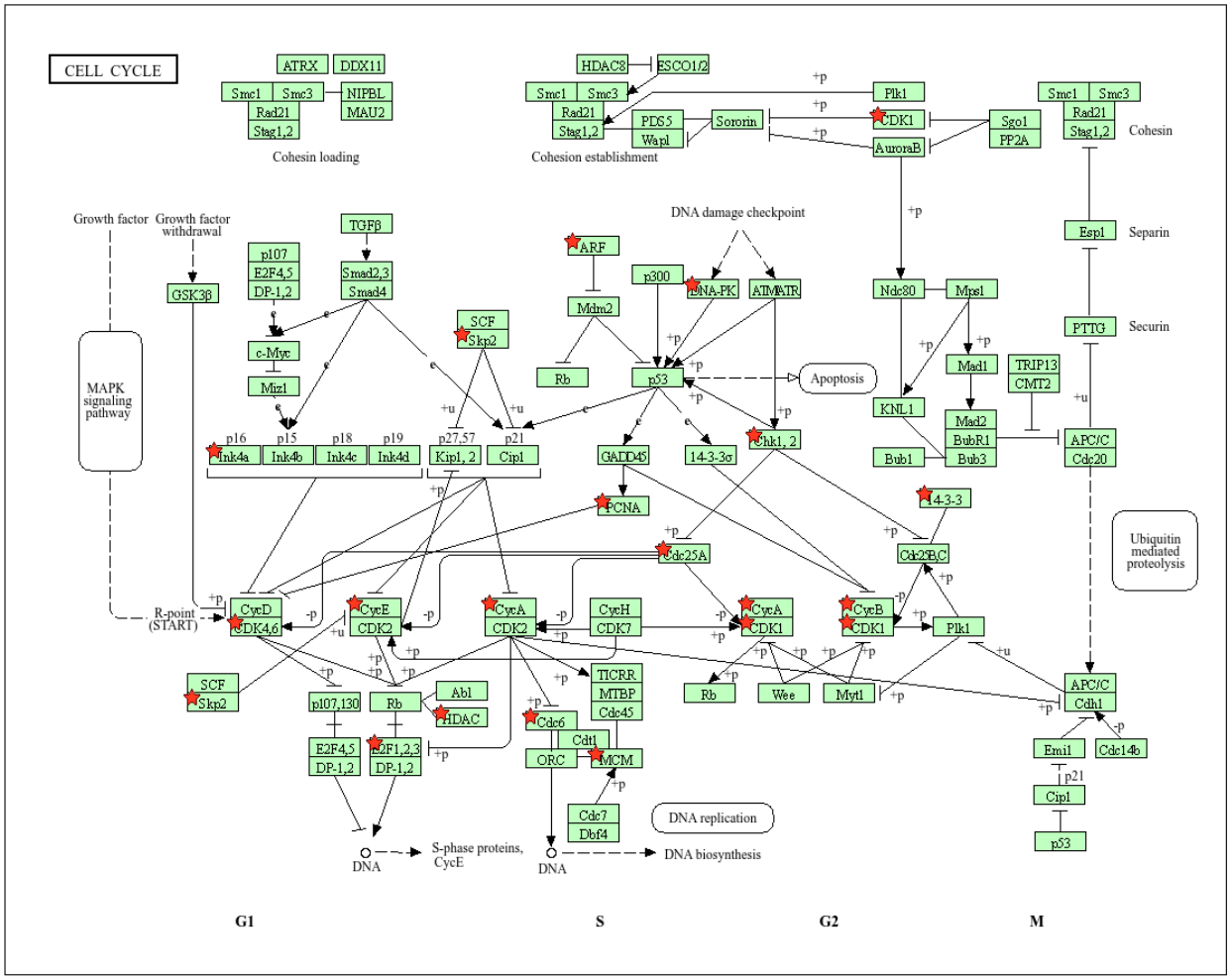


Figure 3: KEGG pathway flow illustrating the Cell Cycle pathway from DAVID, with upregulated genes marked by red stars. This figure highlights the critical components of the cell cycle pathway and emphasizes the genes that are upregulated, which are indicated by red stars. These genes cell cycle progression, contributing to the dysregulation of normal cellular processes in cancer. The red stars indicate key genes within this pathway that may drive tumorigenesis by disrupting normal cell cycle regulation.

Similar to the Cell Cycle pathway, another significantly enriched pathway is the HIF-1 signaling pathway (hsa04066), with genes such as *LDHA, EGLN3, SLC2A1, EIF4EBP1, PGK1, TIMP1, ENO1, ALDOA, GAPDH, PFKP,* and *PDK1* being upregulated (**Figure 4**). The upregulation of these genes suggests a critical role of hypoxia in lung cancer, as these genes are involved in metabolic adaptation to low oxygen conditions. This highlights the importance of hypoxia-induced signaling in promoting tumor survival and progression through altered cellular metabolism.

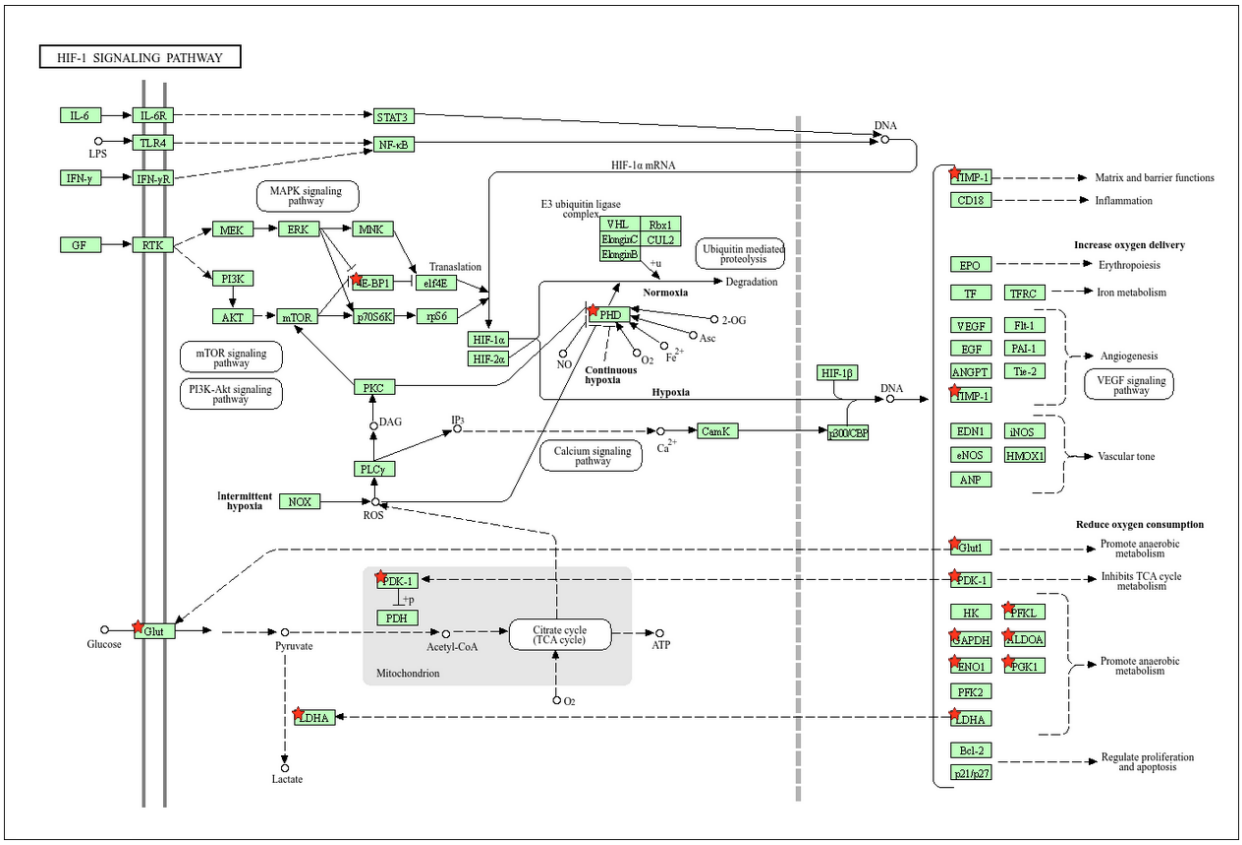


Figure 4: KEGG pathway flow illustrating the HIF-1 signaling pathway from DAVID, with upregulated genes marked by red stars. This figure highlights the enrichment of genes associated with the HIF-1 signaling pathway, which plays a crucial role in cellular responses to hypoxia. Upregulated genes, such as LDHA, EGLN3, SLC2A1, EIF4EBP1, PGK1, TIMP1, ENO1, ALDOA, GAPDH, PFKP, and PDK1, are shown with red stars, indicating

their involvement in metabolic reprogramming and cellular adaptation to low oxygen conditions.

In addition to the upregulated genes, we observed that the downregulated genes are predominantly associated with immune-related pathways and signaling processes. As illustrated in **Figure 5**, several immune-related pathways are significantly affected, indicating a potential suppression of immune responses in lung cancer. Key pathways include the TNF signaling pathway (hsa04668), Cytokine-cytokine receptor interaction (hsa04060), and IL-17 signaling pathway (hsa04657), all of which play essential roles in immune activation and inflammation. Other important pathways such as Toll-like receptor signaling (hsa04620), JAK-STAT signaling (hsa04630), and Natural killer cell-mediated cytotoxicity (hsa04650) are also downregulated, suggesting that the immune system's ability to recognize and eliminate cancer cells may be impaired. The statistical analysis of upregulated pathways is tabulated in **Supplementary Table S6**. These findings suggest a potential immune evasion mechanism in lung cancer, where the downregulation of key immune pathways may allow tumors to escape immune surveillance and progress unchecked.

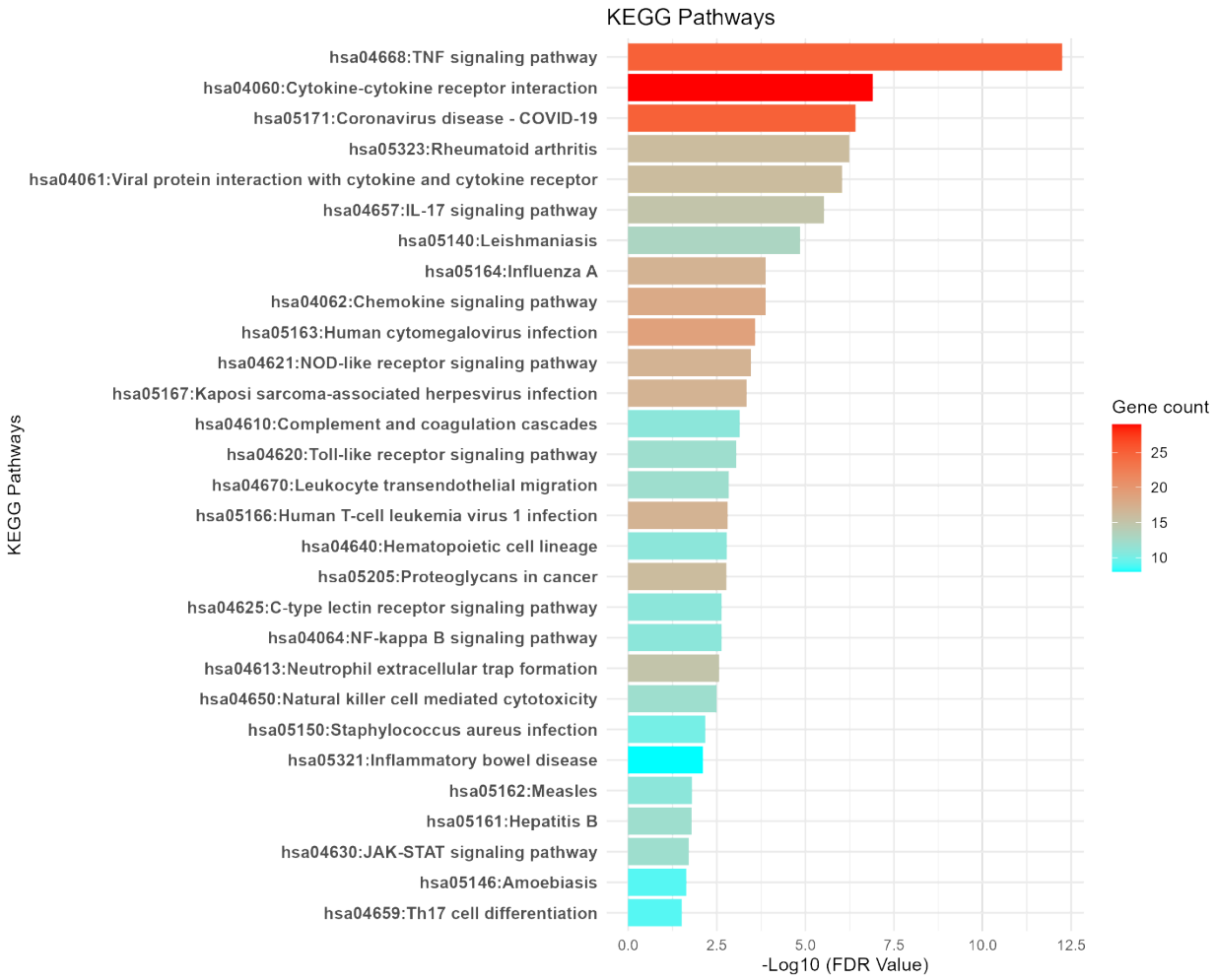


Figure 5: Immune-associated pathways enriched in the downregulated genes. This figure illustrates the major immune-related pathways that are significantly downregulated in lung cancer, suggesting impaired immune responses. Key pathways include the TNF signaling pathway (hsa04668), Cytokine-cytokine receptor interaction (hsa04060), IL-17 signaling pathway (hsa04657), and JAK-STAT signaling pathway (hsa04630), along with several other pathways related to immune cell activation, inflammation, and pathogen recognition.

Explorations of hub HRDEGs

Next, I conducted a Protein-Protein Interaction (PPI) analysis by inputting all the hypoxia-related differentially expressed genes (DEGs) into the STRING database. The primary PPI network consisted of 600 nodes and 9,281 edges, with an average node degree of 30.9 (**Figure 6**). The PPI

enrichment p-value was found to be < 1.0e-16, indicating a highly significant interaction among the hypoxia-related genes. Based on a degree of interaction greater than 10, we identified the hub genes within the network. After identifying the hub genes based on the degree of interaction, we found that the top hub genes include *GAPDH, IL6, TNF, IL1B, MMP9, JUN, CCL2, TLR4, FOS, PTGS2, PECAM1, CDH1, HSP90AB1, PTPRC, ICAM1, FGF2, PPARG, ITGAM, CXCL12, and KDR*. The list of all interacted genes and their degree of interaction provided in the **Supplementary Table S7**. These genes play pivotal roles in key biological processes such as inflammation, immune response, cell signaling, and tumor progression. **Figure 7A** illustrates the interaction network of the top 50 hub genes an, highlighting their interconnections and potential involvement in the molecular pathways that drive lung cancer progression. We further identified 201 common upregulated hub genes and 224 common downregulated hub genes, which were shared across the datasets. These genes represent key molecular players in the tumorigenic processes of lung cancer, with their dysregulation potentially contributing to cancer progression. The results of this analysis are presented in **Figure 7B**, and are also detailed in **Supplementary Table S7** for further reference.

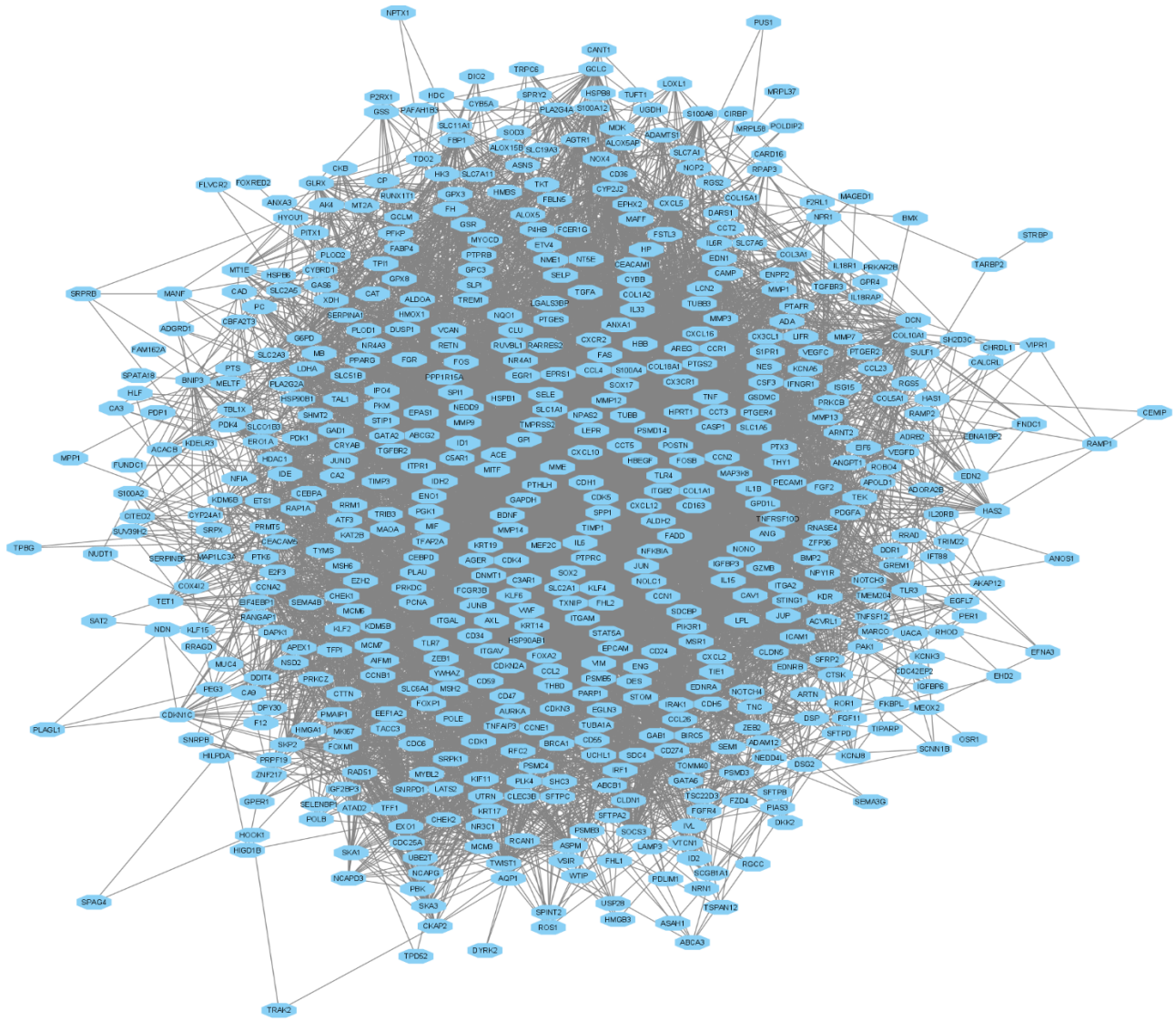


Figure 6. Primary Protein-Protein Interaction (PPI) network of hypoxia-related differentially expressed genes (DEGs) generated using STRING. The network consists of 600 nodes and 9,281 edges, with an average node degree of 30.9. The PPI enrichment p-value is < 1.0e-16, indicating a highly significant interaction between the hypoxia-related genes.

The top upregulated hub genes identified in this analysis include *MMP9, CDH1, HSP90AB1, SOX2, CDKN2A, SPP1, EZH2, TIMP1, CXCL10, COL1A1, HDAC1, BRCA1,* and *CDK4*. These genes are involved in critical processes such as tumor invasion, immune modulation, cell cycle

regulation, and epigenetic changes. *MMP9* and *SPP1* play key roles in extracellular matrix remodeling and metastasis, while *CDH1* is important for maintaining cell-cell adhesion. *SOX2* and *EZH2* regulate stem cell-like properties and epigenetic modifications, contributing to cancer stem cell maintenance and progression. *CDK4* and *CDKN2A* are crucial regulators of the cell cycle, promoting cell division and tumor growth. *BRCA1* and *HDAC*1 are involved in DNA repair and chromatin remodeling, while *CXCL10* and *TIMP1* are associated with immune responses and tissue remodeling. The upregulation of these hub genes highlights their critical roles in tumor progression and their potential as therapeutic targets in lung cancer.

The top downregulated hub genes identified in this analysis *include IL6, TNF, IL1B, JUN, CCL2, TLR4, FOS, PTGS2, PECAM1, PTPRC, ICAM1, FGF2, PPARG, ITGAM, CXCL12,* and *KDR*. These genes are primarily involved in immune response, inflammation, and cellular signaling *pathways. The downregulation of IL6, TNF, and IL1B suggests impaired inflammatory signaling,* while *TLR4, FOS*, and *ICAM1* are involved in immune cell activation and adhesion, indicating potential immune evasion mechanisms. *PECAM1* and *KDR* are associated with endothelial cell function and angiogenesis, both crucial for tumor vasculature. The downregulation of *PPARG*, *ITGAM*, and *CXCL12* further highlights disruptions in immune regulation and cell migration. Overall, the downregulation of these hub genes suggests a weakened immune response, contributing to tumor progression and immune evasion in lung cancer.

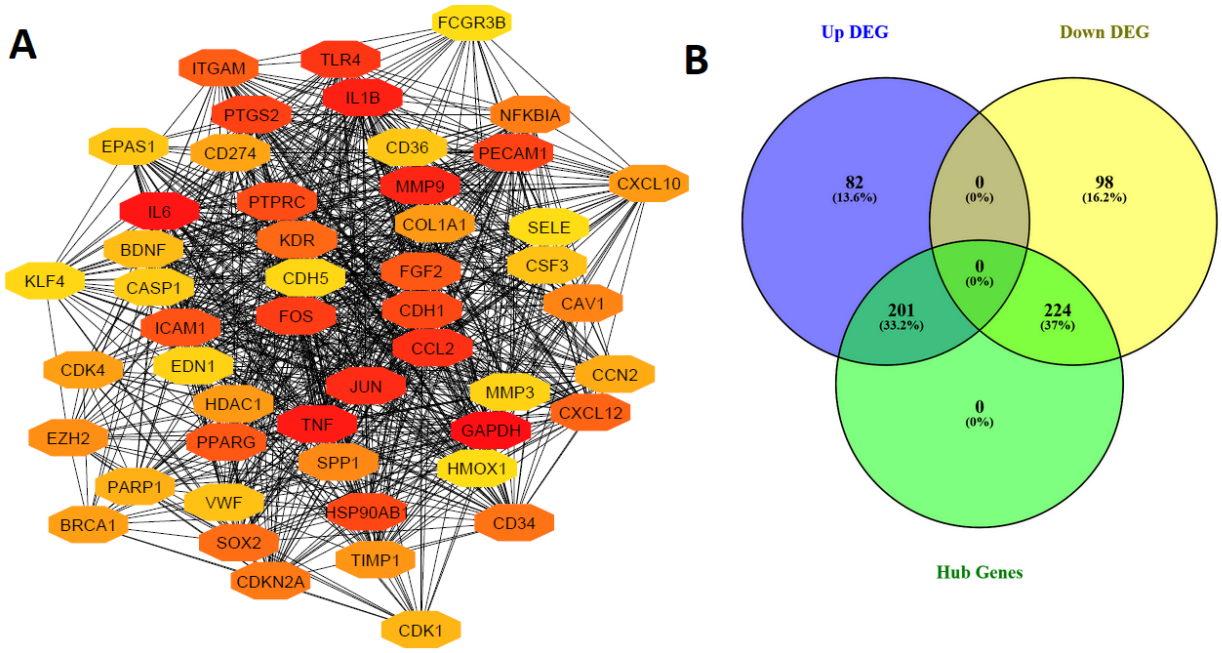


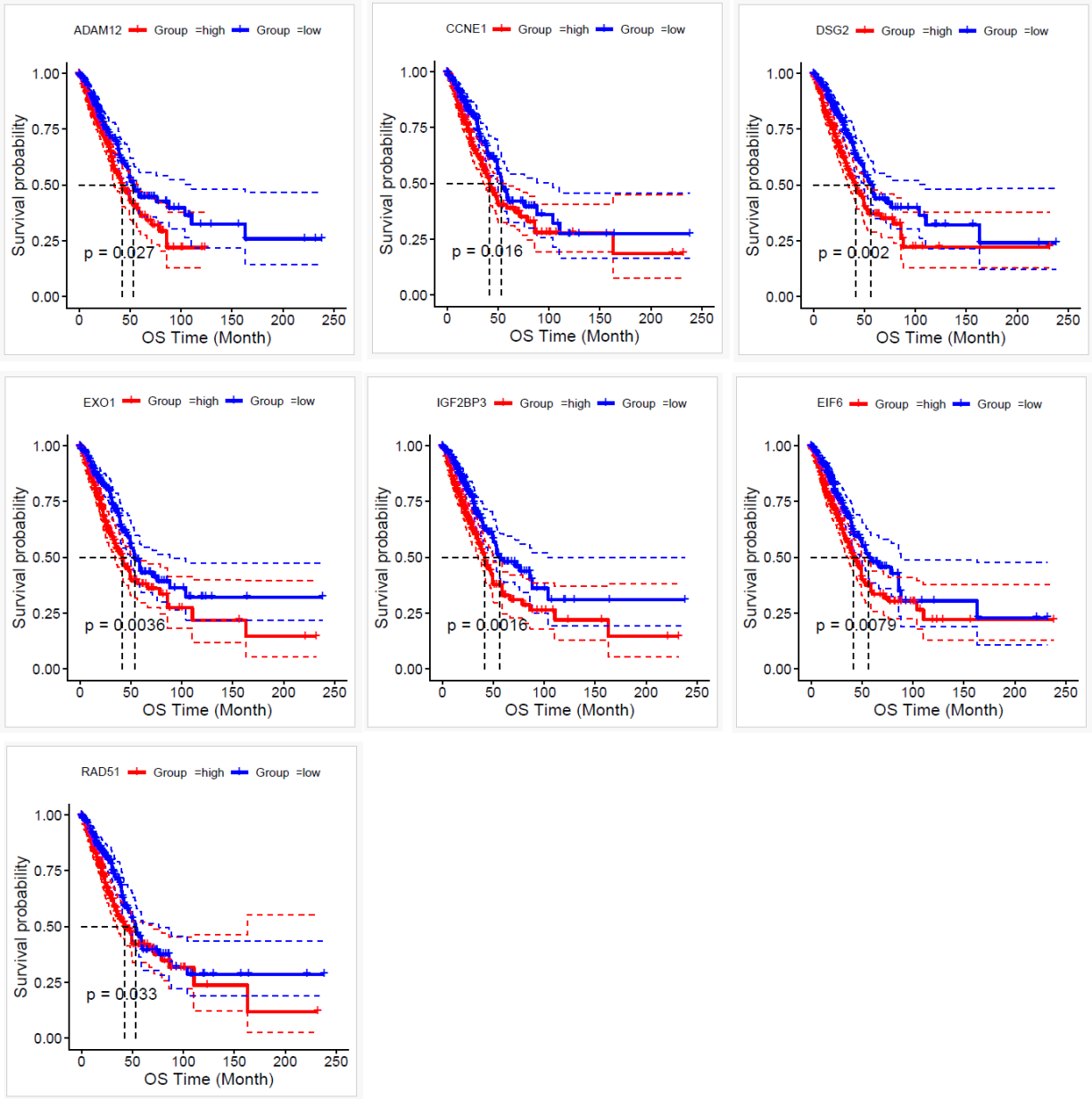
Figure 7. Protein-Protein Interaction (PPI) Network of Hub Genes. (A) Interaction network of the top 50 hub genes identified from the PPI analysis. These genes, with high degrees of interaction, play crucial roles in lung cancer progression, including tumor invasion, immune modulation, and cell cycle regulation. (B) Common hub genes shared between the upregulated and downregulated gene sets.

Exploration of independent prognostic HRDEGs

We first conducted Kaplan-Meier survival analysis on all hub genes and identified 174 significant genes that were associated with survival (**Supplementary Table S8**). The following seven genes (Identified by Univariate regression and LASO)—ADAM12, CCNE1, DSG2, EIF6, EXO1, IGF2BP3, and RAD51—are identified as upregulated hub genes in lung cancer (**Figure 8A**). The upregulated group, categorized based on median expression levels, is associated with worse survival outcomes. Increased expression of these genes is correlated with poorer prognosis,

suggesting that they may play a critical role in tumor progression and could serve as potential biomarkers for identifying high-risk patients with lung cancer. The following ten genes ((Identified by Univariate regression and LASO))—ABCB1, ADRB2, ALDH2, CAT, CLEC3B, FBLN5, MAP3K8, PIK3R1, SFTPD, and SOD3—are identified as downregulated hub genes in lung cancer. The downregulated group, categorized based on median expression levels, is associated with worse survival outcomes. Lower expression of these genes correlates with poorer prognosis, suggesting that their downregulation may contribute to tumor progression and aggressive cancer behavior. These findings highlight the potential role of these downregulated genes in worsening survival outcomes and underscore their relevance as prognostic biomarkers for lung cancer. These genes, selected through Lasso, represent the most influential biomarkers for survival prediction, highlighting their potential roles in prognosis and therapeutic targeting.

A


ADAM12 Group =high Group =low
p = 0.027
CCNE1 Group =high Group =low
p = 0.016
DSG2 Group =high Group =low
p = 0.002
EXO1 Group =high Group =low
p = 0.0036
IGF2BP3 Group =high Group =low
p = 0.0016
EIF6 Group =high Group =low
p = 0.0079
RAD51 Group =high Group =low
p = 0.033
Survival probability
OS Time (Month)

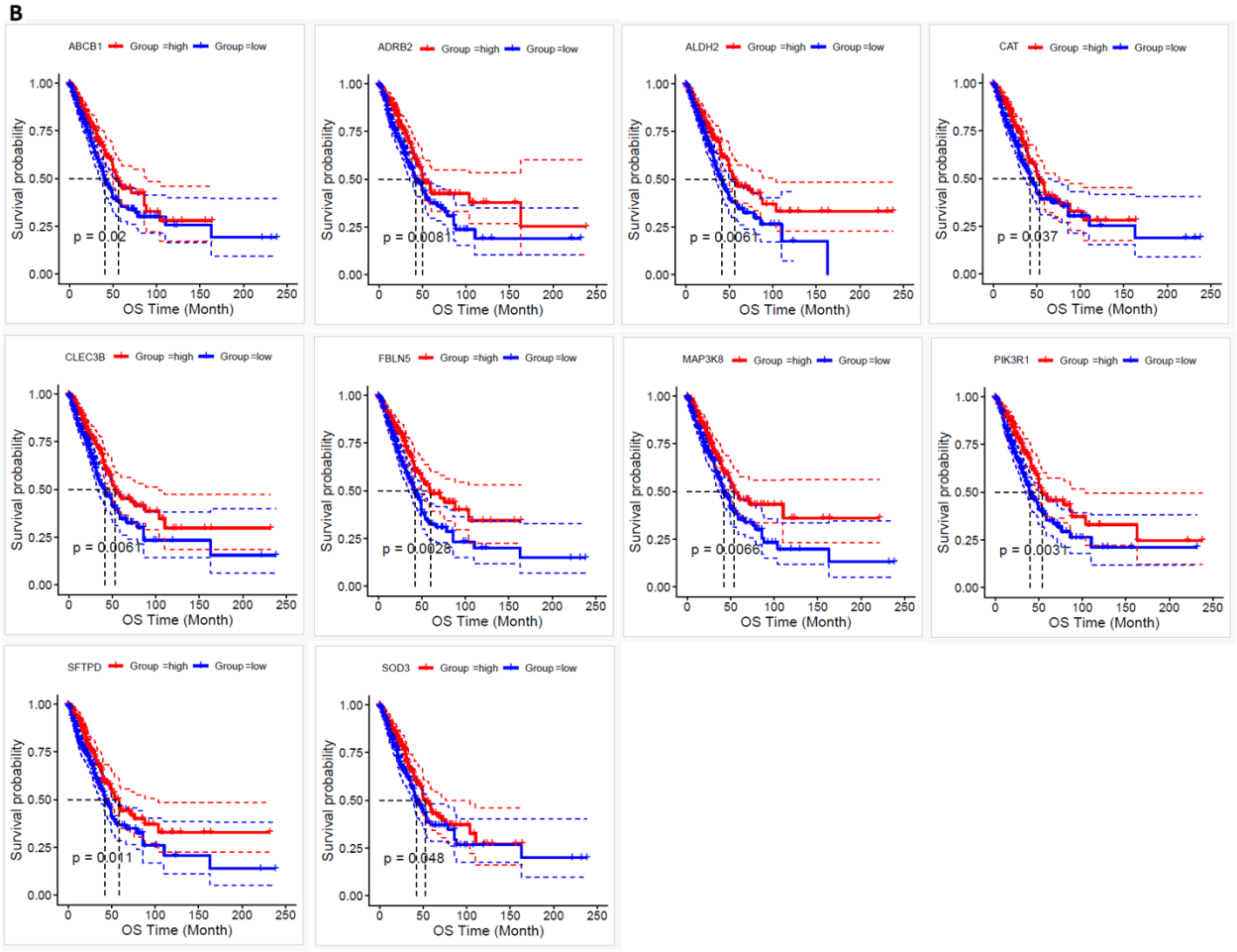


Figure 8. Kaplan-Meier Survival Analysis of Upregulated and Downregulated Hub Genes in Lung Cancer. (A) Kaplan-Meier survival analysis of the group of upregulated hub genes, showing their significant association with survival in lung cancer patients. (B) Kaplan-Meier survival analysis of the group of downregulated hub genes, demonstrating their significant impact on survival outcomes. Both groups reveal distinct prognostic potential, providing insights into the survival-related roles of these gene sets in lung cancer.

To refine this list, we performed univariate regression analysis, which revealed 106 genes with statistically significant associations (**Supplementary Table S9)**. Subsequently, we applied Lasso

regression to the consistently upregulated and downregulated genes to identify key genes with non-zero coefficients. This analysis led to the identification of 17 key genes: *ADRB2, ALDH2, CAT, CCNE1, MAP3K8, DSG2, EIF6, ABCB1, PIK3R1, RAD51, SFTPD, SOD3, CLEC3B, ADAM12, EXO1, FBLN5,* and *IGF2BP3*. The univariate regression analysis and LASSO regression analysis depicted in the **Figure 9** and **Figure 10**, respectively.

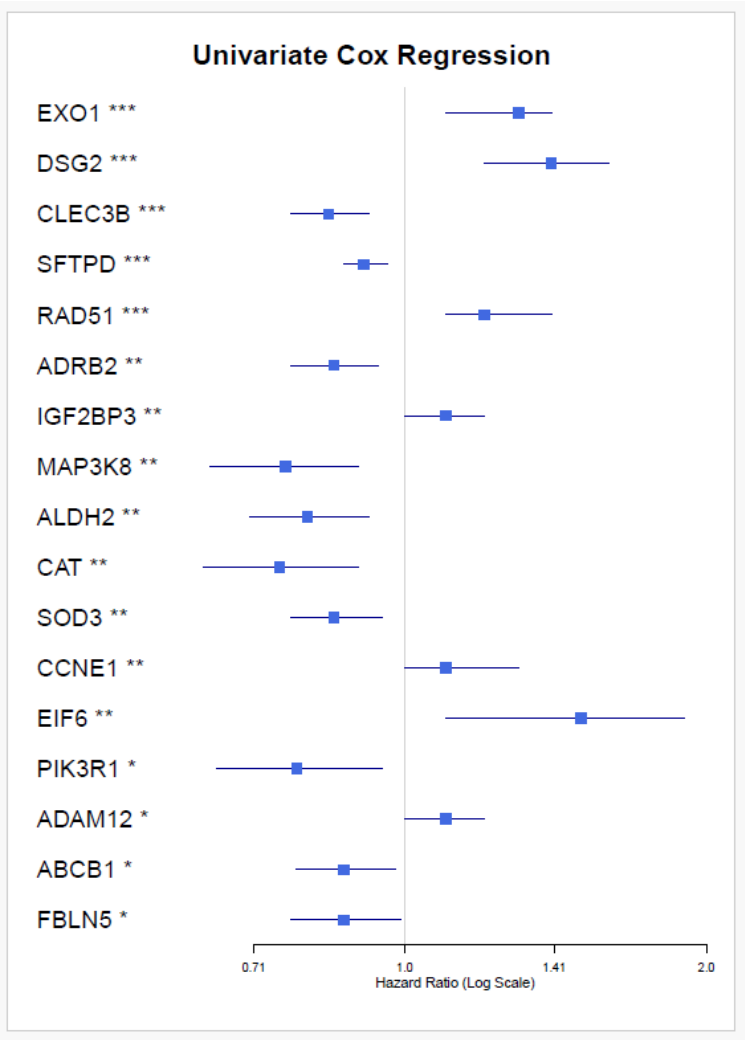


Figure 9: Univariate Cox Regression Analysis. Kaplan-Meier survival curves of hub genes identified through univariate Cox regression analysis. The plot shows the relationship

between the expression levels of significant genes and survival outcomes in lung cancer patients. The survival curves are grouped based on the expression of each gene, and statistical significance is indicated by p-values.

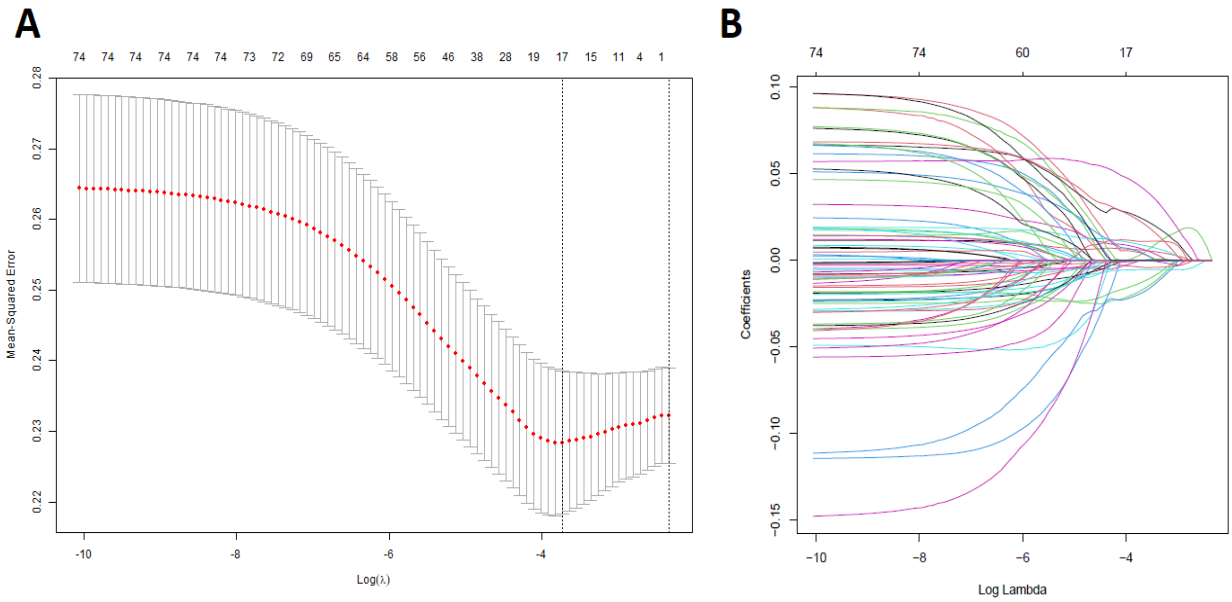


Figure 10: Lasso Regression Analysis of Key Genes. Identification of key genes through Lasso regression. (A) Lasso coefficients for each gene selected through Lasso regression. The plot shows the evolution of the coefficients for each gene as the regularization penalty (lambda) increases, highlighting the key genes that retained non-zero coefficients. (B) Log(lambda) vs. coefficients plot demonstrating the relationship between the log-transformed lambda values and the corresponding coefficients. The plot illustrates how the lambda penalty influences the coefficients of the genes, with non-zero coefficients representing the most influential genes for survival prediction. A total of 17 key genes (*ADRB2, ALDH2, CAT, CCNE1, MAP3K8, DSG2, EIF6, ABCB1, PIK3R1, RAD51, SFTPD, SOD3, CLEC3B, ADAM12, EXO1, FBLN5,* and *IGF2BP3*) were identified, as detailed in the

results, which were then further tested for their potential as prognostic biomarkers for lung cancer survival.

Furthermore, a multivariate Cox regression analysis was conducted to identify independent prognostic factors. The results indicated that *DSG2, EIF6,* and *EXO1* emerged as independent prognostic factors. The multivariate model demonstrated a significant fit, with the following statistics: Concordance = 0.657 (SE = 0.024), Likelihood ratio test = 50.39 on 17 degrees of freedom (df), p = 4e-05, Wald test = 48.47 on 17 df (p = 7e-05), and Score (logrank) test = 48.73 on 17 df (p = 7e-05). *DSG2* (HR = 1.238, p = 0.012), *EIF6* (HR = 1.387, p = 0.046), and *EXO1* (HR = 1.257, p = 0.047) all have hazard ratios greater than 1, indicating that higher expression of these genes is associated with poorer survival outcomes. These results further support the robustness of *DSG2, EIF6,* and *EXO1* as key genes in the prediction of survival, making them critical targets for future therapeutic strategies and prognostic modeling.

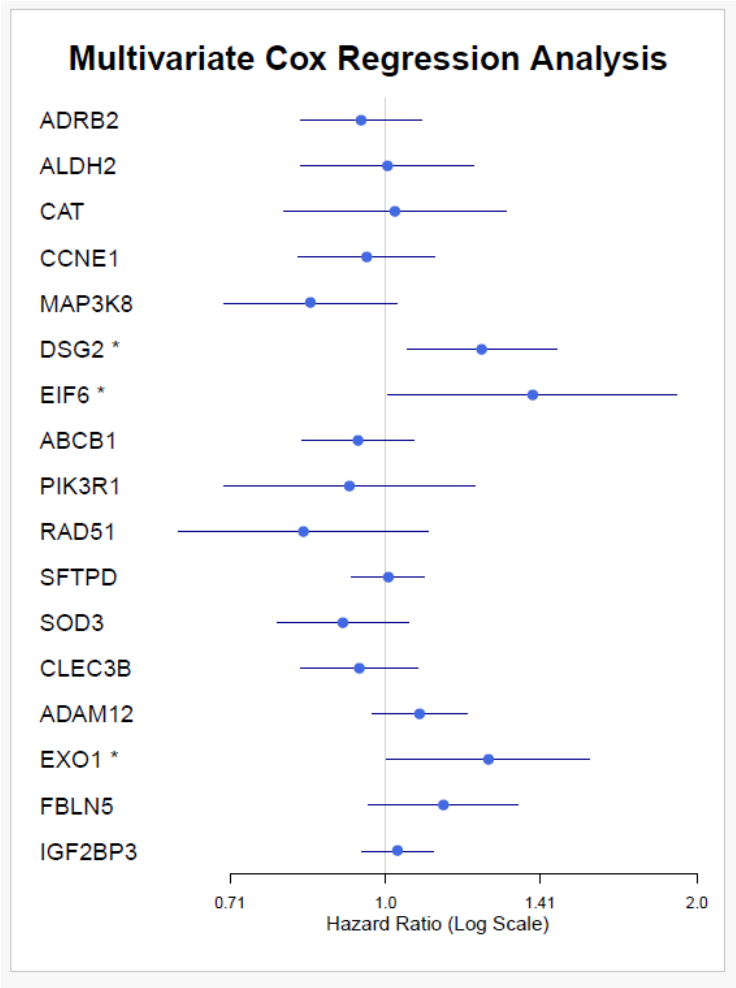


Figure 11: Multivariate Cox Regression Analysis. Multivariate Cox regression analysis identifying independent prognostic factors. This figure presents the results from multivariate Cox regression, highlighting *DSG2, EIF6,* and *EXO1* as independent prognostic factors. The analysis was performed on the 17 key genes identified through Lasso regression. Statistical tests, including the Likelihood Ratio Test, Wald Test, and Score Test, demonstrated that the multivariate model was statistically significant (p <4e-05), indicating that these three genes are key independent predictors of survival in lung cancer patients.

Conclusions

In conclusion, our analysis identifies *DSG2, EIF6,* and *EXO1* as key hypoxia-related genes with significant prognostic value in lung adenocarcinoma. These genes, through their involvement in hypoxia-related pathways, may play critical roles in tumor progression and patient survival. The findings suggest that *DSG2, EIF6,* and *EXO1* could serve as potential biomarkers for predicting patient outcomes and may offer novel therapeutic targets for improving prognosis in lung adenocarcinoma. Further validation of these genes in larger cohorts and clinical settings is necessary to confirm their utility in clinical practice.

Acknowledgements

Work in the SR lab is funded by the Wellcome Trust (206293/Z/17/Z).

References

1. Zhang, Y.; Vaccarella, S.; Morgan, E.; Li, M.; Etxeberria, J.; Chokunonga, E.; Manraj, S.S.; Kamate, B.; Omonisi, A.; Bray, F. Global Variations in Lung Cancer Incidence by Histological Subtype in 2020: A Population-Based Study. *Lancet Oncol.* **2023**, *24*, 1206–1218, doi:10.1016/S1470-2045(23)00444-8.
2. Seguin, L.; Durandy, M.; Feral, C.C. Lung Adenocarcinoma Tumor Origin: A Guide for Personalized Medicine. *Cancers* **2022**, *14*, 1759, doi:10.3390/cancers14071759.
3. Zhang, C.; Liu, Y.; Lu, Y.; Chen, Z.; Liu, Y.; Mao, Q.; Bao, S.; Zhang, G.; Zhang, Y.; Lin, H.; et al. Identification of Potential Biomarkers for Lung Adenocarcinoma: A Study Based on Bioinformatics Analysis Combined with Validation Experiments. *Front. Oncol.* **2024**, *14*, 1425895, doi:10.3389/fonc.2024.1425895.
4. Emami Nejad, A.; Najafgholian, S.; Rostami, A.; Sistani, A.; Shojaeifar, S.; Esparvarinha, M.; Nedaeinia, R.; Haghjooy Javanmard, S.; Taherian, M.; Ahmadlou, M.; et al. The Role of Hypoxia in the Tumor Microenvironment and Development of Cancer Stem Cell: A Novel Approach to Developing Treatment. *Cancer Cell Int.* **2021**, *21*, 62, doi:10.1186/s12935-020-01719-5.
5. Ciepła, J.; Smolarczyk, R. Tumor Hypoxia Unveiled: Insights into Microenvironment, Detection Tools and Emerging Therapies. *Clin. Exp. Med.* **2024**, *24*, 235, doi:10.1007/s10238-024-01501-1.
6. Petrova, V.; Annicchiarico-Petruzzelli, M.; Melino, G.; Amelio, I. The Hypoxic Tumour Microenvironment. *Oncogenesis* **2018**, *7*, 1–13, doi:10.1038/s41389-017-0011-9.
7. Bae, T.; Hallis, S.P.; Kwak, M.-K. Hypoxia, Oxidative Stress, and the Interplay of HIFs and NRF2 Signaling in Cancer. *Exp. Mol. Med.* **2024**, *56*, 501–514, doi:10.1038/s12276-024-01180-8.

8. Bui, B.P.; Nguyen, P.L.; Lee, K.; Cho, J. Hypoxia-Inducible Factor-1: A Novel Therapeutic Target for the Management of Cancer, Drug Resistance, and Cancer-Related Pain. *Cancers* **2022**, *14*, 6054, doi:10.3390/cancers14246054.
9. Gonzalez-Avila, G.; Sommer, B.; Flores-Soto, E.; Aquino-Galvez, A. Hypoxic Effects on Matrix Metalloproteinases' Expression in the Tumor Microenvironment and Therapeutic Perspectives. *Int. J. Mol. Sci.* **2023**, *24*, 16887, doi:10.3390/ijms242316887.
10. Roy, S.; Kumaravel, S.; Sharma, A.; Duran, C.L.; Bayless, K.J.; Chakraborty, S. Hypoxic Tumor Microenvironment: Implications for Cancer Therapy. *Exp. Biol. Med.* **2020**, *245*, 1073–1086, doi:10.1177/1535370220934038.
11. Chen, Z.; Han, F.; Du, Y.; Shi, H.; Zhou, W. Hypoxic Microenvironment in Cancer: Molecular Mechanisms and Therapeutic Interventions. *Signal Transduct. Target. Ther.* **2023**, *8*, 1–23, doi:10.1038/s41392-023-01332-8.
12. Matic, N.; Pettersson, L.; Sellebjerg, F.; Lindberg, L.; Roberg, K.; Wiechec, E. Prognostic Value of Hypoxia-Responsive Gene Expression Profile in Patients Diagnosed with Head and Neck Squamous Cell Carcinoma. *Transl. Oncol.* **2023**, *39*, 101841, doi:10.1016/j.tranon.2023.101841.
13. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* **2009**, *37*, 1–13, doi:10.1093/nar/gkn923.
14. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300.
15. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613, doi:10.1093/nar/gky1131.
16. Chin, C.-H.; Chen, S.-H.; Wu, H.-H.; Ho, C.-W.; Ko, M.-T.; Lin, C.-Y. cytoHubba: Identifying Hub Objects and Sub-Networks from Complex Interactome. *BMC Syst. Biol.* **2014**, *8*, S11, doi:10.1186/1752-0509-8-S4-S11.
17. Therneau, T. A Package for Survival Analysis in R. 95.
18. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22.

Supplementary Information

Supplementary Tables S1-S9

Supplementary Table S1. List of collected hypoxia related genes.

Supplementary Table S2. List of DEGs in TCGA LUAD.

Supplementary Table S3. List of DEGs in GSE18842.

Supplementary Table S4. List of hypoxia related DEGs in the both datasets.

Supplementary Table S5. Enriched pathways associated with the upregulated genes (FDR < 0.05).

Supplementary Table S6. Enriched pathways associated with the downregulated genes (FDR < 0.05).

Supplementary Table S7. Rank of the PPI network by Degree method.

Supplementary Table S8. Significant Genes Identified in Kaplan-Meier Survival Analysis.

Supplementary Table S9. Significant Genes Identified in Univariate regression Analysis.