Integrated bioinformatics and feature selection-based discovery of ADP-ribosylation-associated biomarkers in non-small cell lung cancer

JIE HE[1], QINGTAO ZHAO[1,2], HONGZHEN ZHAO[1,2], HUA ZHANG[1], ZHONGHUI HU[1], LINGXIN KONG[1], ZENGMING WANG[1], JIANGHONG YAN[3], SHUO YANG[3] and SHICHENG LIU[1,2]

[1]Department of Thoracic Surgery, Hebei General Hospital, Shijiazhuang, Hebei 050057, P.R. China;
[2]Graduate School, Hebei Medical University, Shijiazhuang, Hebei 050017, P.R. China;
[3]Department of Pediatric Research, Children's Hospital of Hebei Province, Shijiazhuang, Hebei 050000, P.R. China

Received November 8, 2026; Accepted May 1, 2026

DOI: 10.3892/ol.2026.15852

Abstract. Non-small cell lung cancer (NSCLC) remains a significant global health concern. ADP-ribosylation, a key post-translational modification, may influence the tumor immune microenvironment; however, the genes transcriptionally linked to ADP-ribosylation pathway activity in NSCLC are poorly understood. In the present study, transcriptomic data from the Gene Expression Omnibus database were analyzed to identify differentially expressed genes (DEGs) associated with ADP-ribosylation via differential expression analysis. Weighted gene co-expression network analysis was subsequently applied to refine these DEGs and extract ADP-ribosylation-related module genes. Feature selection methods aided by machine learning, including LASSO regression and the Boruta algorithm, were employed to identify core signature genes. Enzyme-linked immunosorbent assay (ELISA) validated poly(ADP-ribose) polymerase (PARP) protein levels in NSCLC tissues. Immune infiltration analysis, single-gene gene-set enrichment analysis and regulatory network construction were used to explore gene functions. ELISA confirmed significantly higher PARP concentrations in NSCLC tissues compared to para-cancerous tissues (P<0.01), indicating dysregulated ADP-ribosylation pathway activity. Three signature genes (HBZ, HLA-DRA and CCL4) were identified, all of which were significantly associated with immune-cell infiltration and immune factor expression. Reverse transcription-quantitative PCR validation in 10 paired NSCLC patient samples revealed a significant downregulation of HBZ (P<0.0001) and CCL4 (P<0.01), while HLA-DRA showed a non-significant downward trend. This study provides preliminary evidence that HBZ, HLA-DRA and CCL4 are transcriptionally linked to ADP-ribosylation pathway activity and may modulate the NSCLC immune microenvironment, offering potential avenues for future immunotherapeutic research.

Correspondence to: Dr Shicheng Liu, Department of Thoracic Surgery, Hebei General Hospital, 348 Heping West Road, Xinhua, Shijiazhuang, Hebei 050057, P.R. China
E-mail: 1309388955@qq.com

Key words: non-small cell lung cancer, ADP-ribosylation, immune microenvironment, feature selection, biomarkers, poly(ADP-ribose) polymerase, enzyme-linked immunosorbent assay

Introduction

According to statistics from the World Health Organization, the annual incidence of non-small cell lung cancer (NSCLC) exceeds 2.1 million new cases, with nearly 1.5 million deaths, making its high morbidity and mortality a major global public health challenge (1). In China, the 5-year survival rate for patients with NSCLC remains <30%, with most cases diagnosed at advanced stages (2,3). As the predominant histological subtype, NSCLC accounts for ~85% of all lung cancer cases, with lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) being its principal subtypes (4). LUSC, which constitutes ~30% of NSCLC cases, has a higher propensity for metastasis and recurrence, making it particularly aggressive and therapeutically challenging with a poor prognosis (4-6). Advancements in precision medicine strategies have become central to addressing these challenges, with ongoing research on new molecular targets offering hope for improved treatment options. Notably, the identification of EGFR as a target has significantly increased survival rates for patients with NSCLC (7,8).

ADP-ribosylation, a post-translational modification, involves the transfer of ADP-ribose moieties to target proteins catalyzed by ADP-ribosyltransferases. This process is pivotal in key biological functions, including DNA damage repair, gene expression regulation, cell cycle progression and signal transduction (9,10). Aberrant ADP-ribosylation has been closely linked to tumorigenesis and cancer progression. Dysregulated activation of poly(ADP-ribose) polymerase 1 (PARP1) can lead to increased genomic instability, promoting tumor cell proliferation and survival (11,12). The tumor microenvironment (TME) consists of tumor cells, immune

cells (such as macrophages), cytokines and other components (13,14). The heterogeneity within the TME plays a pivotal role in facilitating immune evasion in NSCLC (14). For instance, elevated soluble CD47 in the serum of patients with lung cancer inhibits the phagocytic activity of alveolar macrophages, contributing to an immunosuppressive environment (15). Although immune checkpoint blockade therapy has significantly improved outcomes for certain patients, only a subset of patients benefit from this treatment (16). Therefore, identifying novel immunotherapeutic targets remains of paramount clinical importance.

ADP-ribosylation may influence immune-cell infiltration and function by modifying immune-related proteins or regulating the secretion of inflammatory cytokines (17,18). However, its specific role in the immune microenvironment of NSCLC remains inadequately explored. Current research primarily focuses on individual genes or pathways, with limited integrative analyses of ADP-ribosylation-related genes in relation to immune cell infiltration and cytokine expression. Furthermore, the immune regulatory functions of ADP-ribosylation-associated genes in NSCLC have not been fully examined in existing databases.

This study aimed to identify and validate genes transcriptionally associated with ADP-ribosylation pathway activity that may participate in immune regulation through multi-omics integrative analysis. To provide direct protein-level evidence for ADP-ribosylation pathway dysregulation, PARP protein levels in NSCLC and para-cancerous tissues were quantified using enzyme-linked immunosorbent assay (ELISA). A total of three ADP-ribosylation signature genes (HBZ, HLA-DRA and CCL4) were identified and their roles in NSCLC were preliminarily investigated through multi-omics analyses. Notably, the genes identified by bioinformatics methods were selected based on their transcriptomic correlation with ADP-ribosylation pathway signatures, rather than direct measurements of ADP-ribosylation activity. These findings offer novel insights into ADP-ribosylation-mediated immune regulation and provide a theoretical foundation for the development of new immunotherapeutic strategies.

Materials and methods

Data resources and identification of DEGs. Gene expression data for NSCLC and the ADP-ribosylation-related gene collection were retrieved from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). Data from the GSE68465 dataset were utilized in conjunction with the Molecular Signatures Database (MSigDB) under the BIOCARTA_ARAP_PATHWAY module (19). The GSE68465 dataset comprises 442 NSCLC samples (including both LUAD and LUSC subtypes) and 20 normal lung tissue samples. Detailed clinical annotations, including histological subtypes and disease stage, are available in the GEO repository. Differential gene expression profiles between NSCLC tumor samples and their corresponding normal tissue counterparts were analyzed using the limma package in R software (version 4.4.2; R Foundation for Statistical Computing; https://www.r-project.org/). DEGs were identified based on stringent statistical criteria: |log2-fold change|>2 and P<0.05, ensuring the selected genes represented significant biological differences. A total of 754 DEGs were identified after thorough analysis. Limma-based differential expression analysis was performed on the normalized expression matrix from GEO, with no additional batch-effect correction applied due to the single-dataset design.

ELISA for PARP protein quantification. To provide direct protein-level evidence for the dysregulation of ADP-ribosylation pathway activity in NSCLC, ELISA was performed to measure PARP protein concentrations in NSCLC and adjacent para-cancerous tissues. A total of 8 paired NSCLC tumor tissues and matched para-cancerous tissues were collected from patients who underwent surgical resection at Hebei General Hospital (Shijiazhuang, China). Tissue samples were homogenized in ice-cold radioimmunoprecipitation assay lysis buffer supplemented with protease inhibitors, and total protein was extracted by centrifugation at 13,000 x g for 15 min at 4˚C. Protein concentrations were determined using a bicinchoninic acid protein assay kit. PARP protein levels were quantified using a commercially available human PARP ELISA kit following the manufacturer's instructions (cat. no. MK1220HB; Human Poly(ADP-Ribose) ELISA Research Kit; 48 Tests). In brief, tissue lysates were added to pre-coated 96-well microplates and incubated at 37˚C for 2 h. After washing, the detection antibody was added and incubated for 1 h at 37˚C, followed by incubation with HRP-conjugated secondary antibody for 30 min. 3,3',5,5'-tetramethylbenzidine substrate solution was added for color development and the reaction was terminated with stop solution. The optical density was measured at 450 nm using a microplate reader. PARP concentrations (pg/ml) were calculated based on a standard curve. Each sample was assayed in duplicate wells. Data are presented as mean ± SD. Statistical comparison between groups was performed using the paired Student's t-test, with $P<0.05$ considered statistically significant.

Weighted gene co-expression network analysis (WGCNA). The expression data matrix was further analyzed using the R package WGCNA, as described in a previous study (20). Hierarchical clustering of all datasets was performed using the Euclidean distance metric between gene expression levels to identify and exclude outlier samples. Prior to constructing the co-expression network, the optimal soft threshold value was determined to ensure that the resulting network exhibited scale-free topological characteristics. During the WGCNA analysis, the correlation coefficient weighted value was used to establish gene connections that followed a scale-free distribution. A soft threshold power ranging from 1 to 30 was selected to define the criterion for gene correlation. An adjacency matrix was then constructed and converted into a topological overlap matrix. Finally, a hierarchical clustering approach was applied to classify genes into distinct functional groups (gene modules), with these clusters visually represented through color coding for enhanced interpretability.

This study used the single-sample gene set enrichment analysis (ssGSEA) scores of ADP-ribosylation genes as phenotypes to explore associations between phenotypes and modules and calculate the correlation coefficient matrix between module eigengenes (MEs) and phenotypic traits. Notably, ADP-ribosylation activity was not directly measured at the

transcriptomic level in this study. Instead, the ssGSEA scores derived from the BIOCARTA_ARAP_PATHWAY gene set (obtained from the Molecular Signatures Database, MSigDB v2024.1.Hs; https://www.gsea-msigdb.org/) served as a transcriptomic surrogate. ssGSEA scoring was performed using the GSVA R package (version 2.0.7). Consequently, the genes identified should be interpreted as transcriptionally correlated with this pathway signature. Complementary protein-level validation of ADP-ribosylation activity was provided by measuring PARP concentrations through ELISA. Modules were considered trait-associated if they met the criteria of Pearson correlation coefficient $|r| > 0.3$ and $P < 0.05$, and were selected for further analysis. Genes from both modules that exhibited the strongest positive and negative correlations with phenotypes were filtered using the criteria module membership (MM) >0.7 and gene significance (GS) >0.2, resulting in 103 genes. These genes were considered as ADP-ribosylation-related module genes for downstream analysis.

Candidate genes identification and protein-protein interaction (PPI) network construction. Next, the 754 DEGs were intersected with the 103 ADP-ribosylation-related module genes. The overlapping genes, referred to as 'ADP-ribosylation-related DEGs', were designated as candidate genes for simplicity. Through computational screening, a total of 69 potential genetic markers were identified. A comprehensive PPI network for these 69 candidate genes was constructed using the Search Tool for the Retrieval of Interacting Genes and proteins (STRING) database (http://string-db.org). For network inclusion, an interaction score cutoff of >0.4 was applied and Cytoscape software (version 3.9.1; https://cytoscape.org/) was used to visualize the interconnected nodes.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. GO and KEGG pathway enrichment analyses were performed on the 69 candidate genes using the 'clusterProfiler' package in R. These analyses aimed to identify significant biological pathways associated with these genes, with statistical significance evaluated at the $P < 0.05$ threshold.

Machine learning analysis. To further identify critical genes among the 69 candidates, two complementary machine learning-based feature selection methods were applied: The least absolute shrinkage and selection operator (LASSO) and the Boruta algorithm, a robust feature selection technique. LASSO regression and Boruta analysis were conducted using the glmnet and Boruta packages in R, respectively. It is important to note that LASSO and Boruta were used solely for feature selection and not for predictive model construction. No classification model, training/testing split or performance metrics [e.g., area under the curve (AUC), accuracy] were generated in this study. Genes selected by both LASSO and Boruta were designated as feature genes, resulting in a total of 3 feature genes for subsequent analyses. To evaluate the diagnostic discriminative capacity of the identified feature genes, an independent receiver operating characteristic (ROC) curve analysis was performed in the TCGA LUAD subset of GSE68465 using the pROC R package. Area under the curve (AUC) and 95% confidence intervals were computed for each individual gene and for the combined three-gene signature represented by the first principal component (PC1).

Single-gene GSEA. To elucidate the biological functions of the three feature genes-HBZ, HLA-DRA and CCL4-the 'c5.go.bp.v2024.1.Hs.symbols' gene set from the MSigDB database was selected as the reference. Spearman correlation analysis was performed to determine the association between the feature genes and all other genes in the dataset. Based on these correlation coefficients, a single-gene GSEA was conducted on the feature genes using the 'clusterProfiler' package in R, allowing exploration of the functional implications of their expression.

Analysis of immune infiltration and differentially expressed immune factors. To estimate the proportions of 22 immune cell types in all samples, the R package 'CIBERSORT' was utilized. Spearman correlation analyses between the feature genes and differentially abundant immune cells across all samples were performed using the 'psych' package in R. Expression levels of immune factors-24 immune checkpoint inhibitors, 45 immune stimulators and 41 chemokines-were compared using the Wilcoxon test. Immune factors with significant differential expression were categorized as 'differentially expressed immune factors' for downstream analysis. A network diagram was then constructed using Cytoscape to visualize interactions among the feature genes, differentially expressed immune factors and differentially abundant immune cells.

Feature gene-centered regulatory network establishment. To identify microRNA (miRNA) candidates that may regulate the feature genes, three data repositories-miRWALK (http://mirwalk.umm.uni-heidelberg.de) (21), miRDB (http://www.mirdb.org) (22) and TargetScan (http://www.targetscan.org) (23)-were employed. High-confidence miRNAs were defined as those consistently identified across all three platforms. Additionally, the KnockTF database (http://www.licpathway.net/KnockTF/index.html) (24) was used to identify transcription factors (TFs) potentially involved in regulating HBZ, HLA-DRA and CCL4. The integrated miRNA-TF regulatory network for these feature genes was constructed using the miRNet tool (https://www.mirnet.ca/) and visualized with Cytoscape software (v3.9.1) (25).

RNA extraction and reverse transcription-quantitative (RT-q) PCR analysis. A total of 10 pairs of NSCLC and adjacent normal tissues were collected from NSCLC patients who underwent surgical treatment at the Department of Thoracic Surgery of Hebei General Hospital between March and June 2024. The detailed clinical information of all enrolled patients and tissue samples is presented in Table SI. All patients were pathologically diagnosed with NSCLC. Inclusion criteria were as follows: i) Histopathologically confirmed primary NSCLC; ii) curative-intent surgery without prior neoadjuvant chemotherapy, radiotherapy or immunotherapy; and iii) availability of paired tumor and adjacent (≥5 cm from tumor edge) non-tumor tissues. Exclusion criteria were as follows: i) Other concurrent malignancies; ii) preoperative antitumor treatment; and iii) inadequate tissue quality for RNA extraction. After surgical excision, the collected tissues were rinsed with

ice-cold PBS. Total RNA was extracted using a chloroform-based method with a commercial RNA extraction reagent (cat. no. G3013; Wuhan Servicebio Technology Co., Ltd.) according to the manufacturer's instructions. RNA concentration and purity were determined using a Nanodrop 2000 (Thermo Fisher Scientific, Inc.).

cDNA synthesis was performed using a reverse transcription kit (cat. no. G3337; Wuhan Servicebio Technology Co., Ltd.) in a 20-μl reaction volume, containing 4 μl of 5X SweScript All-in-One SuperMix for qPCR, 1 μl of gDNA Remover, 10 μl of total RNA, and nuclease-free water to adjust the final volume. The mixture was gently mixed, briefly centrifuged and subjected to reverse transcription on a conventional PCR instrument under the following conditions: 25˚C for 5 min, 42˚C for 30 min and 85˚C for 5 sec. For qPCR, reactions were prepared in triplicate in a 0.1 ml RT-qPCR plate. Each 15 μl reaction consisted of 7.5 μl of 2X Universal Blue SYBR Green qRT-PCR Master Mix, 1.5 μl of 2.5 μM gene-specific primers (forward + reverse), 2.0 μl of cDNA template and 4.0 μl of nuclease-free water. After loading, the plate was sealed with PCR film, briefly centrifuged and amplified on a real-time PCR instrument using the following protocol: Pre-denaturation: 95˚C for 30 sec. Amplification (40 cycles): 95˚C for 15 sec (denaturation) and 60˚C for 30 sec (annealing/extension). The mRNA expression of the target gene was quantified using the $2^{-\Delta\Delta Cq}$ formula (26). Statistical significance was evaluated using paired Student's t-test (GraphPad Prism 8.0; Dotmatics). All reactions were performed in technical triplicate and results were considered statistically significant at $P<0.05$. Target gene expression was normalized to GAPDH and quantified using the $2^{-\Delta\Delta Cq}$ method (26). Data are presented as mean±standard deviation (SD). Sense (S, forward) and antisense (A, reverse) primer sequences were as follows: H-GAPDH-S, GGAAGCTTGTCATCAATGGAAATC; H-GAPDH-A, TGATGACCCTTTTGGCTCCC; H-HBZ-S, ACTTCAAGCTCCTGTCCCAC; H-HBZ-A, TCAGCG GTACTTCTCGGTCA; H-HLA-DRA-S, CCATAAGTG GAGTCCCTGTGCTA; H-HLA-DRA-A, CAAAGTCAA ACATAAACTCGCCTG; H-CCL4-S, GTAGCTGCCTTC TGCTCTCC and H-CCL4-A, GAAGCTTCCTCGCGG TGTAA.

Results

Identification of DEGs in NSCLC and validation of ADP-ribosylation pathway activity. To identify genes associated with NSCLC, the transcriptome sequencing dataset GSE68465 (442 NSCLC samples vs. 20 normal controls) was retrieved from the GEO database. The workflow used in this study is outlined in Fig. S1. Differential expression analysis was performed to compare the normal and NSCLC groups. The selection criteria for DEGs were as follows: P-value <0.05 and |log2 fold change|>2. A total of 754 DEGs were identified, consisting of 453 upregulated and 301 downregulated genes. The expression patterns of the top 10 genes with the highest fold changes are presented in Fig. 1A; notably, several of these genes are involved in immune regulation and metabolic processes related to lung cancer biology. Heatmaps were generated to visualize the top five upregulated (SFTPB, IGHA2, LOC100509457, HLA-DRA, RGS1) and downregulated (HBZ, HBE1, ALB, AHSG, APOA2) genes in NSCLC (Fig. 1B).

To verify the dysregulation of ADP-ribosylation pathway activity at the protein level, ELISA was performed to measure PARP concentrations in paired NSCLC and adjacent para-cancerous tissues (n=8 per group). As shown in Fig. 1C, PARP protein concentrations were significantly elevated in NSCLC tissues compared to adjacent para-cancerous tissues (~5,000 vs. 4,300 pg/ml, $P<0.01$), confirming that ADP-ribosylation-related enzymatic activity is dysregulated in NSCLC. This provides direct protein-level evidence supporting the biological relevance of the subsequent transcriptomic analysis of ADP-ribosylation-associated genes.

Defining ADP-ribosylation-related module genes. To identify co-expressed gene modules and explore their potential associations with ADP-ribosylation-related transcriptomic traits, WGCNA was employed. Using the ssGSEA scores of ADP-ribosylation as traits, modules statistically correlated with ADP-ribosylation activity ($|r|>0.3$, $P<0.05$) were identified. Hierarchical clustering was first performed to detect and exclude potential outliers; however, no outliers were found (Fig. S2A). To create a scale-free network, an optimal soft threshold power (β) ranging from 1 to 30 was selected. This value was determined through a statistical evaluation of the scale-free topology model's fit, which required an R-squared coefficient of at least 0.85 (Fig. S2B). The mean connectivity values are shown in Fig. S2C, leading to the final selection of $\beta=5$. Genes were then assigned to co-expression modules using this threshold, with each module represented by a unique color (Fig. 2A). A total of six modules (excluding the grey module, which contains unclassified genes) were identified. Correlation analysis between MEs and ADP-ribosylation ssGSEA scores revealed trait-associated modules. The brown module, positively correlated with ADP-ribosylation ($r=0.38$, $P<0.05$), and the red module, negatively correlated ($r=-0.46$, $P<0.05$), were selected for further analysis (Fig. 2B). The brown module contained 1,137 genes (Fig. 2C) and the red module contained 278 genes (Fig. 2D). Key genes within these modules were filtered based on the following criteria: MM >0.7 and GS >0.2, resulting in 103 ADP-ribosylation-related module genes for downstream analysis.

Identification of ADP-ribosylation-related DEGs and PPI construction. A total of 754 DEGs and 103 ADP-ribosylation-related module genes were identified. The intersection of these two gene sets yielded 69 ADP-ribosylation-related DEGs (Fig. 3A), which were designated as 'candidate genes' for further analysis. To explore PPIs among these 69 candidate genes, this study constructed a PPI network using the STRING database with an interaction score threshold of >0.4 and visualized the network using Cytoscape. As shown in Fig. 3B, immune-related genes such as CD4, CD163 and ITGAM were prominently featured, suggesting potential involvement of these candidate genes in immune regulation. Functional annotation of the 69 candidate genes was performed via GO and KEGG pathway analyses ($P<0.05$). GO analysis revealed enrichment in biological processes such as 'leukocyte-mediated immunity' and regulation of immune response via cell

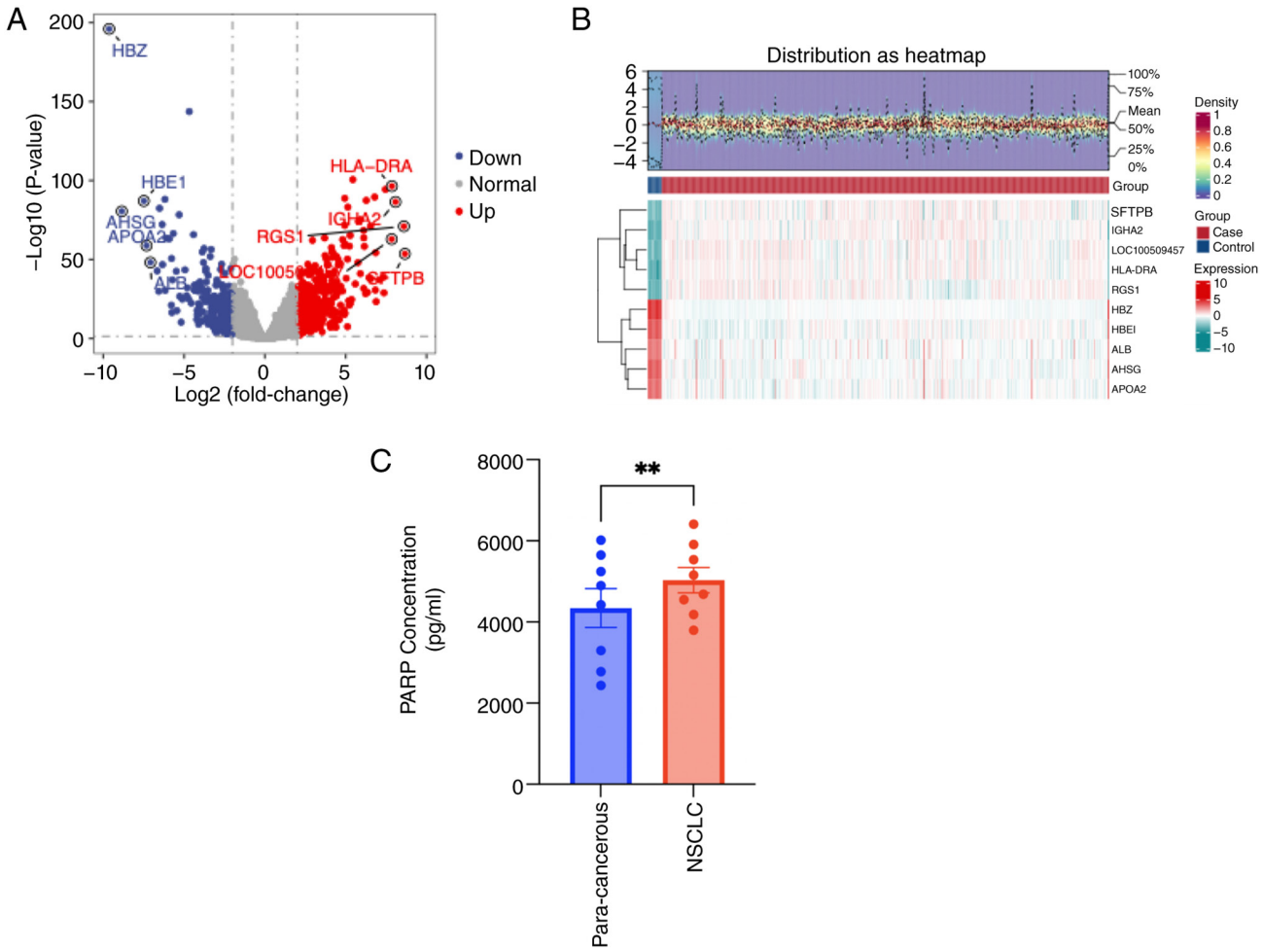

Figure 1. Identification of DEGs. (A) Volcano plot showing the distribution of DEGs, with red indicating upregulated genes, blue indicating downregulated genes and grey indicating genes excluded by DEG screening criteria. (B) Heatmap of DEGs, with the upper part displaying a density heatmap of expression levels for upregulated and downregulated genes across samples, showing lines for five quantiles and the mean. The lower part of the heatmap represents individual samples, highlighting the top 5 upregulated and top 5 downregulated genes. (C) Comparison of PARP concentrations between para-cancerous and NSCLC groups. DEG, differentially expressed gene; NSCLC, non-small cell lung cancer. **P<0.01

surface receptors (Fig. 3C). KEGG pathway analysis revealed significant enrichment in pathways related to host immune defense and innate immune effector functions, with the most significantly enriched terms including '*Staphylococcus aureus* infection', 'Osteoclast differentiation', 'Complement and coagulation cascades' and 'Phagosome' (Fig. 3D).

Machine learning identifies feature genes in NSCLC and ADP-ribosylation. To identify signature genes influencing NSCLC progression via ADP-ribosylation regulation, two machine learning-based feature selection methods were sequentially applied. First, LASSO regression, a regularization technique that introduces a penalty term to the regression model, was employed. This method effectively simplifies the model by driving certain coefficients to zero, enabling variable selection while addressing issues such as high dimensionality, multicollinearity and overfitting in multivariable analyses. LASSO analysis of the 69 candidate genes identified an optimal lambda.min value of 0.006 (Fig. 4A), resulting in three genes with non-zero coefficients: HBZ, HLA-DRA and CCL4 (Fig. 4B). Subsequently, the Boruta algorithm, a wrapper-based feature selection method, was applied to the same set of 69 candidate genes using the training dataset. Boruta iteratively evaluates feature importance by creating shadow features and comparing the importance of real features to the best shadow feature. As shown in Fig. 4C, 33 candidate genes were accepted by the Boruta algorithm. By identifying the overlapping genes selected through both LASSO regression and Boruta feature selection, three consensus signature genes, HBZ, HLA-DRA and CCL4, were determined (Fig. 4D). These genes were prioritized for downstream analyses and functional characterization. To evaluate the diagnostic utility of the three identified signature genes, an independent receiver operating characteristic (ROC) analysis was performed using the TCGA LUAD cohort (GSE68465). The individual gene AUC values were as follows: HLA-DRA (AUC=0.805), CCL4 (AUC=0.633) and HBZ (AUC=0.632). The combined three-gene signature, represented by the first principal component (PC1) of the three genes, achieved an AUC of 0.772 (95% CI: 0.721-0.823), demonstrating that the feature selection process identified biologically meaningful and diagnostically relevant genes with discriminative capacity in an independent subtype-specific cohort (Fig. S3).

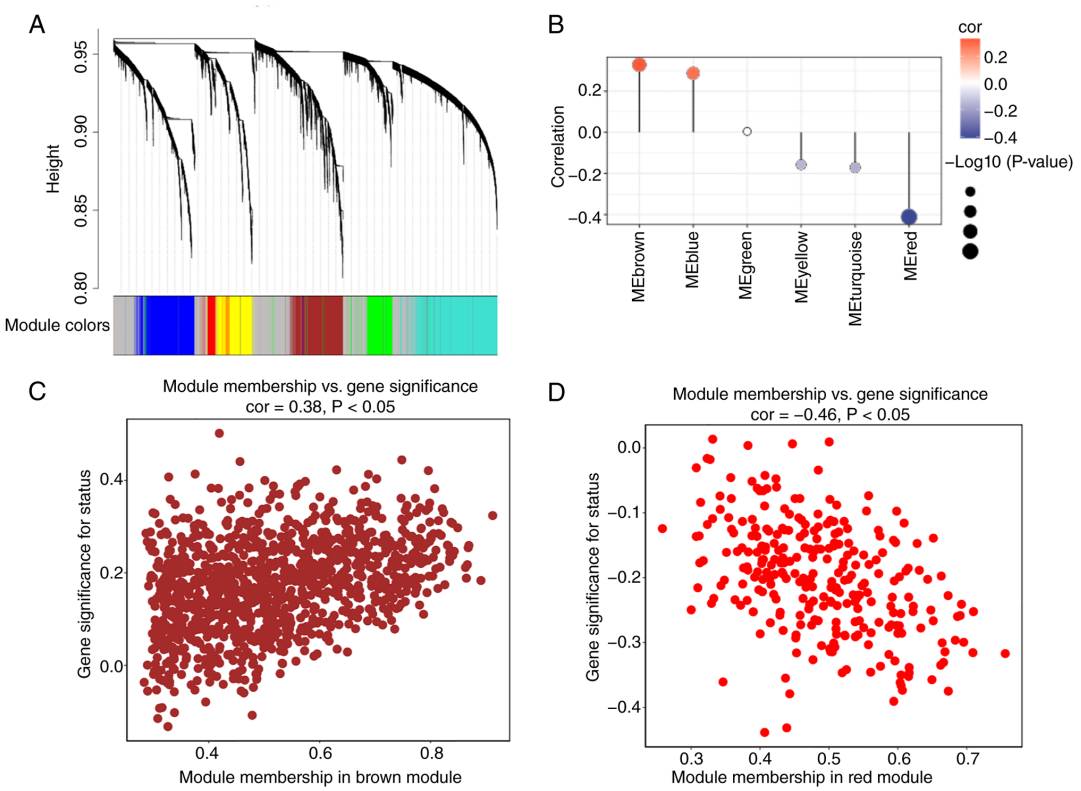

Figure 2. Identification of ADP-ribosylation-related gene modules. (A) Hierarchical clustering dendrogram showing the identification of co-expression modules in genes, with modules represented in the lower part. (B) Lollipop plot illustrating the correlation between modules and phenotypes. The horizontal axis represents modules and the vertical axis represents correlation. Red indicates positive correlation, blue indicates negative correlation, and the size of the circle represents statistical significance. (C and D) Correlation between ADP-ribosylation and (C) the brown module or (D) the red module. Module membership represents the correlation between genes and modules, while gene significance represents the correlation between genes and traits. ME, module eigengene; cor, Pearson correlation coefficient.

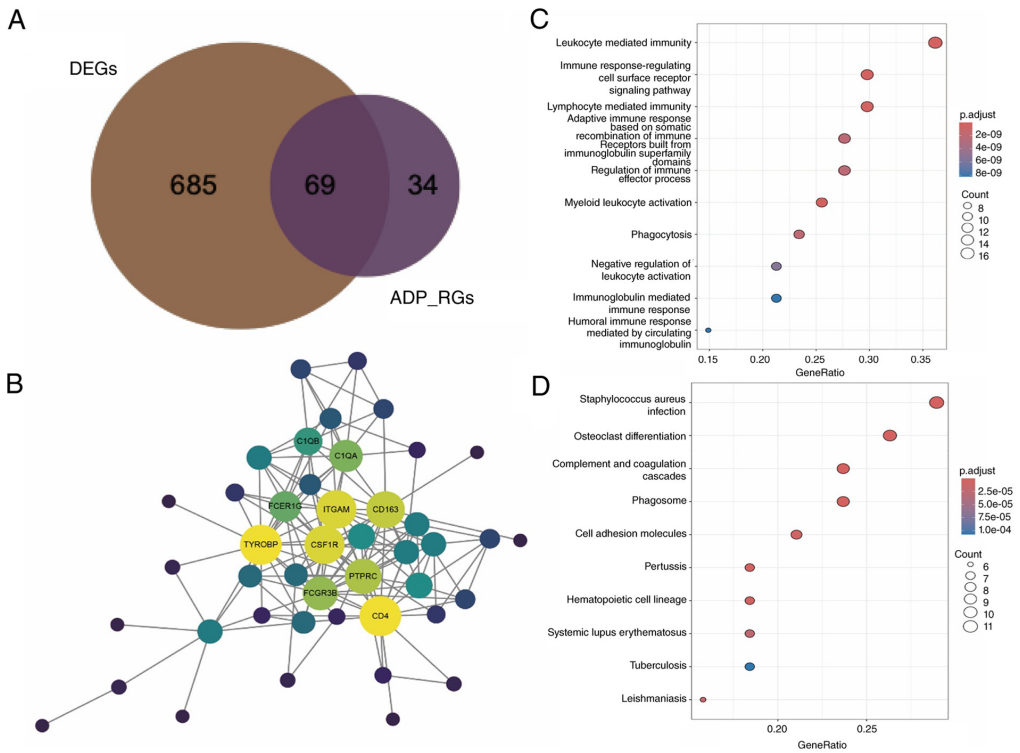

Figure 3. Identification of candidate genes and functional enrichment. (A) Venn diagram showing the intersection between DEGs and ADP-ribosylation-related genes. (B) Protein-protein interaction network of candidate genes, with node color and size representing network centrality. This network highlights the top 10 genes based on degree. (C) Functional enrichment analysis and (D) pathway enrichment analysis for the candidate genes. DEG, differentially expressed gene; ADP_RGs, ADP-ribosylation-related genes; p.Adjust, adjusted P-value.

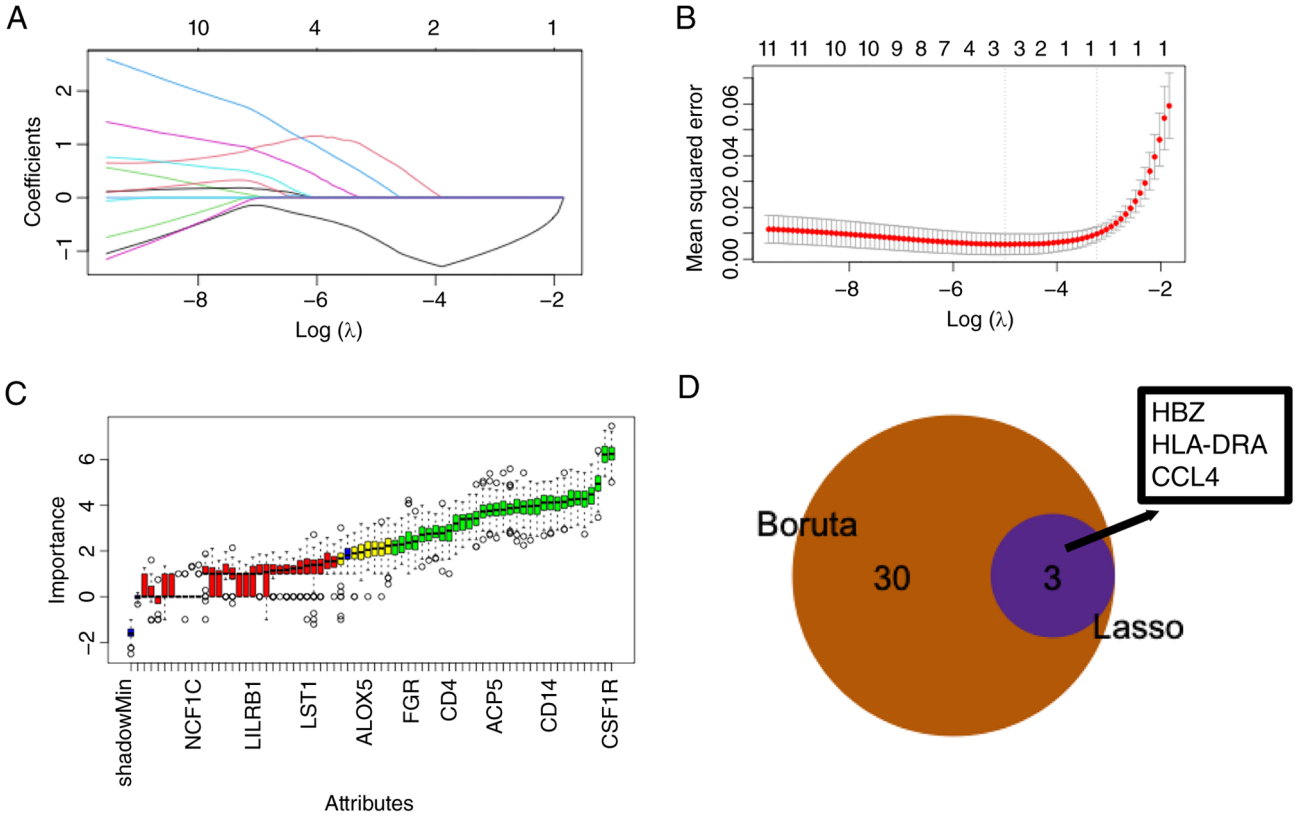

Figure 4. Machine learning and feature gene identification. (A) Coefficient spectrum plot for Lasso model construction, where the x-axis represents the logarithm of lambda values and the y-axis represents variable coefficients. Variables with coefficients equal to 0 are eliminated at the optimal lambda value. (B) 10-fold cross-validation plot for parameter adjustment in the Lasso model, with the x-axis representing the logarithm of lambda values and the y-axis representing the model error. The optimal lambda corresponds to the lowest point of the red curve and selects one variable. (C) Distribution of importance for candidate genes. Red indicates rejected genes, yellow indicates retained genes, and green indicates accepted genes. The vertical axis shows the distribution of importance after multiple cross-validations. (D) Venn diagram showing the genes selected by both the Lasso and Boruta models. Lasso, least absolute shrinkage and selection operator.

Single-gene GSEA analysis of feature genes. To explore the biological functions of the feature genes, single-gene GSEA was performed using the Spearman correlation coefficients between each feature gene and all other genes in the dataset GSE68465. The 'c5.go.bp.v2024.1.Hs.symbols' gene set from the MSigDB database was used as a reference, with correlation coefficients employed as the ranking metric for GSEA. CCL4 positively regulated cellular pathways such as 'peptide antigen assembly with MHC class II protein complex' and negatively regulated the 'regulation of sister chromatid cohesion' pathway (Fig. 5A). HBZ promoted functions such as 'DNA chain elongation during DNA replication' while inhibiting the 'peptide antigen assembly with MHC class II protein complex' pathway (Fig. 5B). HLA-DRA enhanced processes like 'antigen processing and presentation by dendritic cells' and suppressed 'double-strand break repair via break-induced replication' (Fig. 5C).

Immune regulatory network of feature genes. Given the strong association between the three identified feature genes and immune regulation, alterations in the proportions of 22 immune cell subtypes infiltrating NSCLC tissues were investigated. As shown in Fig. 6A and B, significant differences in 18 immune cell types were observed when comparing NSCLC samples with normal lung tissues. Notably altered populations included memory B cells, activated dendritic cells, M0 macrophages, $CD8^+$ T cells, activated natural killer cells and regulatory T cells (Tregs). To further explore the relationship between the three feature genes and these differentially abundant immune cells, Spearman correlation analyses were performed across all samples. The results revealed that each of the 18 differentially represented immune cell types correlated significantly with at least one of the feature genes (Fig. 6C). For instance, CD4 naïve T cells were strongly associated with HBZ expression, while both $CD8^+$ T cells and resting mast cells exhibited robust correlations with HLA-DRA. Similarly, CCL4 expression was significantly linked to M2 macrophages and neutrophils. These results suggest that the altered expression of these feature genes is closely related to changes in the immune microenvironment. To further evaluate immune microenvironment differences within the dataset, the Wilcoxon test was applied to compare the expression levels of immune factors, including 24 immune inhibitors, 45 immune stimulants and 41 chemokines (8), between the NSCLC and normal groups ($P<0.05$). As shown in Fig. 7A, several immune factors, including members of the CCL and CXCL chemokine families, were differentially expressed between NSCLC and normal tissues. To assess the specific pathways through which the feature genes influence the immune microenvironment, a network diagram was constructed using Cytoscape, incorporating the feature genes, differential immune factors and differentially abundant immune cells to illustrate their

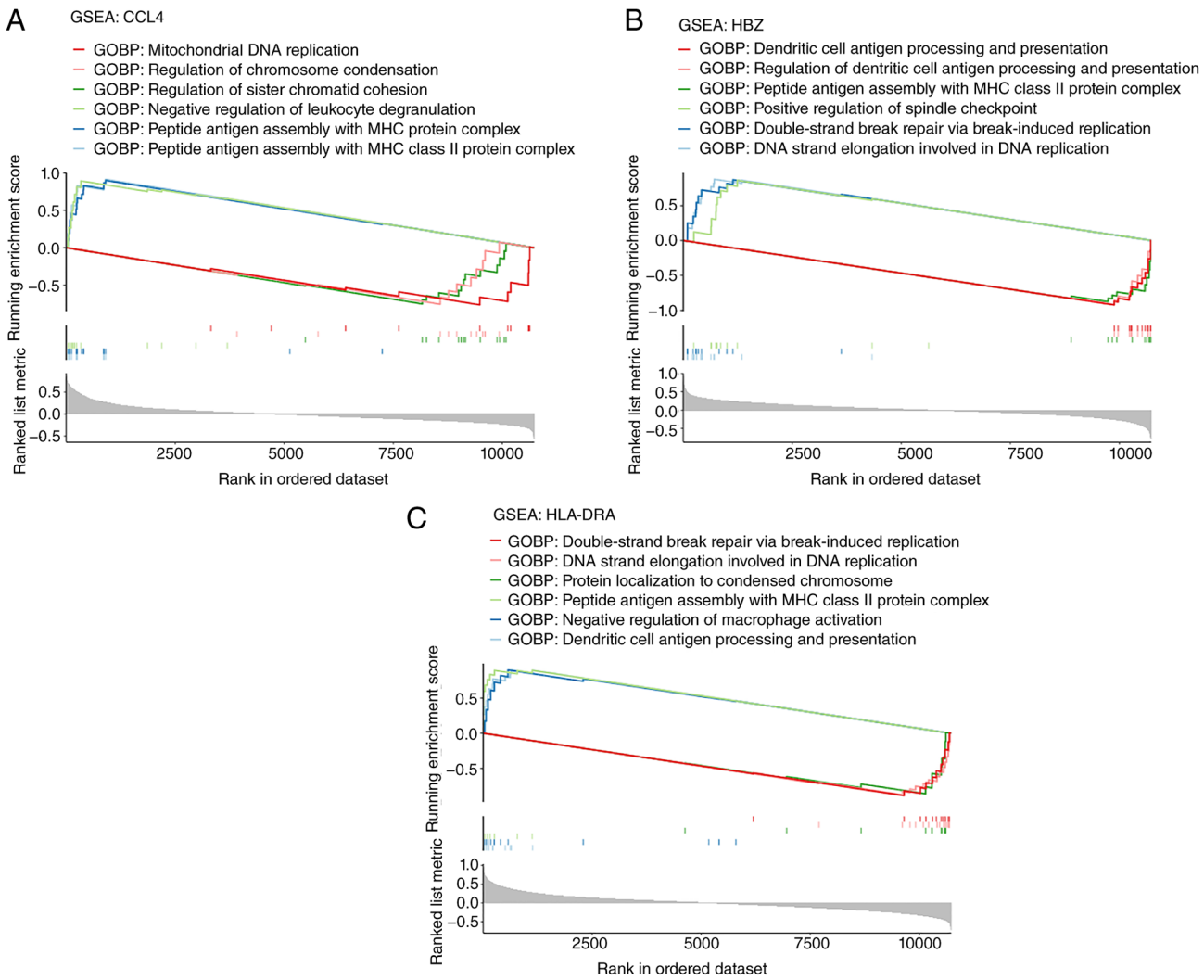

Figure 5. Single-gene GSEA of feature genes. Single-gene GSEA for (A) CCL4, (B) HBZ and (C) HLA-DRA. The upper part shows the functional score curve, the middle part displays the genes involved in the functional pathway and the lower part indicates the correlation between the genes. Genes ranked from highest (left) to lowest (right) correlation with the feature gene; positive enrichment scores indicate functions positively correlated with feature gene expression, and negative scores indicate negative correlation. GSEA, gene-set enrichment analysis; GOBP, Gene Ontology Biological Process.

interactions. The results revealed that the feature genes directly interact with immune cells; for example, the HBZ gene was found to interact directly with CD4 naïve T cells. More commonly, however, the feature genes influence immune cells through cytokines. For instance, CCL4 associates with $CD8^+$ T cells through CCL5. Among the cytokines, the CCL and CXCL families were identified as key mediators facilitating the connection between feature genes and immune cells (Fig. 7B).

Transcriptional regulation and validation expression of feature genes. To explore the molecular regulatory mechanisms underlying the expression of the three identified feature genes, the potential upstream transcriptional regulators were further investigated. Both miRNAs and TFs play critical roles in maintaining cellular homeostasis by modulating the expression of downstream target genes. Therefore, the miRNet database was utilized to model the upstream miRNA-TF regulatory network for these three feature genes, with visualization in Cytoscape. miRNet analysis predicted 91 upstream regulators and 100 regulatory relationships (13 TFs and 78 miRNAs; Fig. 7C). The results revealed that HLA-DRA had the highest number of regulatory relationships. Notably, hsa-miR-26a-5p was found to regulate all three feature genes simultaneously. To validate the significance of the feature genes in NSCLC, tumor and adjacent non-tumor tissues from 10 NSCLC patients were collected and RT-qPCR was performed to assess the expression levels of the three signature genes. As shown in Fig. 7D, both HBZ and CCL4 were significantly downregulated in NSCLC tissues compared to para-cancerous tissues (HBZ: $P<0.0001$; CCL4: $P<0.01$), while HLA-DRA exhibited a downward trend that did not reach statistical significance. The significant downregulation of HBZ and CCL4 provides experimental validation that aligns with the bioinformatics predictions, supporting their roles as ADP-ribosylation-associated biomarkers in NSCLC. The non-significant result for HLA-DRA may be attributable to several factors: First, HLA-DRA expression is known to exhibit considerable inter-individual variability, influenced by various immunological and epigenetic factors. Second, HLA-DRA may primarily exert its functional role through post-transcriptional regulation or by modulating intercellular

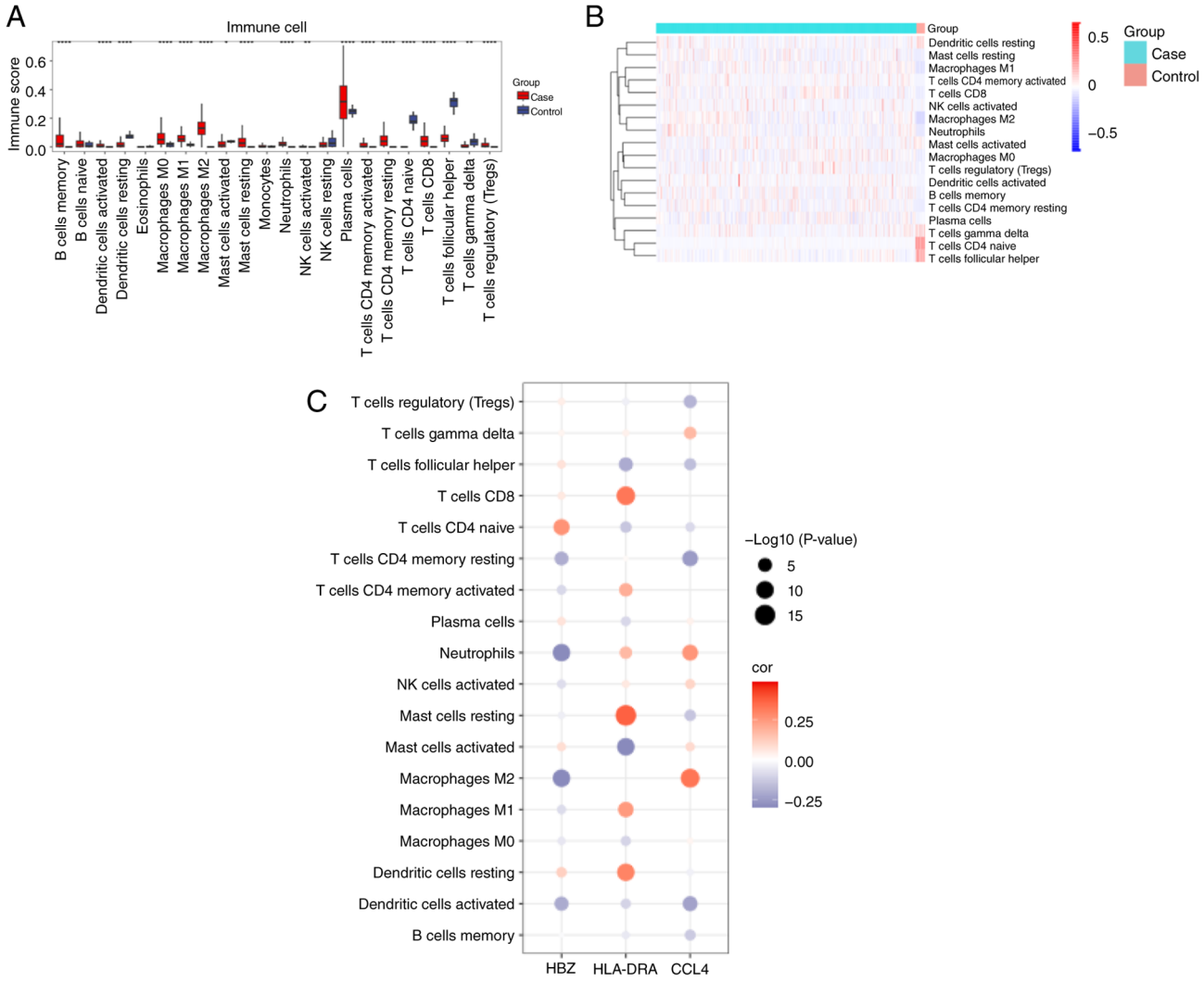

Figure 6. Immune infiltration analysis. (A) Boxplot and (B) heatmap illustrating the distribution of differential immune cell proportions. (C) Correlation between feature genes and differential immune cell proportions. The correlation and significance between the expression of feature genes and differential immune cell scores were calculated. Color represents correlation strength, and the size of the dots represents significance. Non-significant correlations are not displayed. *P<0.05, **P<0.01 and ****P<0.0001. cor, correlation coefficient.

immune communication via MHC class II-mediated antigen presentation, rather than through changes in bulk mRNA expression levels. Third, tumor heterogeneity within NSCLC may obscure expression differences at the tissue level.

Discussion

As the predominant subtype of lung cancer, NSCLC poses significant clinical challenges due to its complex TME and molecular heterogeneity (13,14). Through integrated bioinformatics analysis and experimental validation, this study systematically identified genes transcriptionally associated with ADP-ribosylation pathway activity in NSCLC and preliminarily elucidated their biological functions and clinical relevance. Notably, ELISA revealed significantly elevated PARP protein concentrations in NSCLC tissues compared to para-cancerous tissues (P<0.01), providing direct protein-level evidence for the dysregulation of ADP-ribosylation pathway activity and reinforcing the biological foundation of the present transcriptomic analysis.

The current findings indicate that ADP-ribosylation-associated core signature genes are transcriptionally linked to key biological processes involved in NSCLC progression. Functional enrichment analysis highlights significant involvement of these genes in processes such as MHC class II antigen presentation, DNA damage repair and cell cycle regulation. A comparative analysis between NSCLC and normal lung tissues identified significant differences in the abundance of 18 immune cell types, including Tregs, $CD8^{+}$ T cells and M0 macrophages. By combining differential gene expression analysis, WGCNA, and machine learning-based feature selection, this study identified HBZ, HLA-DRA and CCL4 as core ADP-ribosylation-related signature genes in NSCLC for the first time, to the best of our knowledge. The differential expression of HBZ and CCL4 in NSCLC tissues was further validated by RT-qPCR in an expanded cohort of 10 paired samples, with HBZ showing highly significant downregulation (P<0.0001) and CCL4 demonstrating statistical significance (P<0.01). Additionally, PPI analysis suggested that these signature genes may directly interact

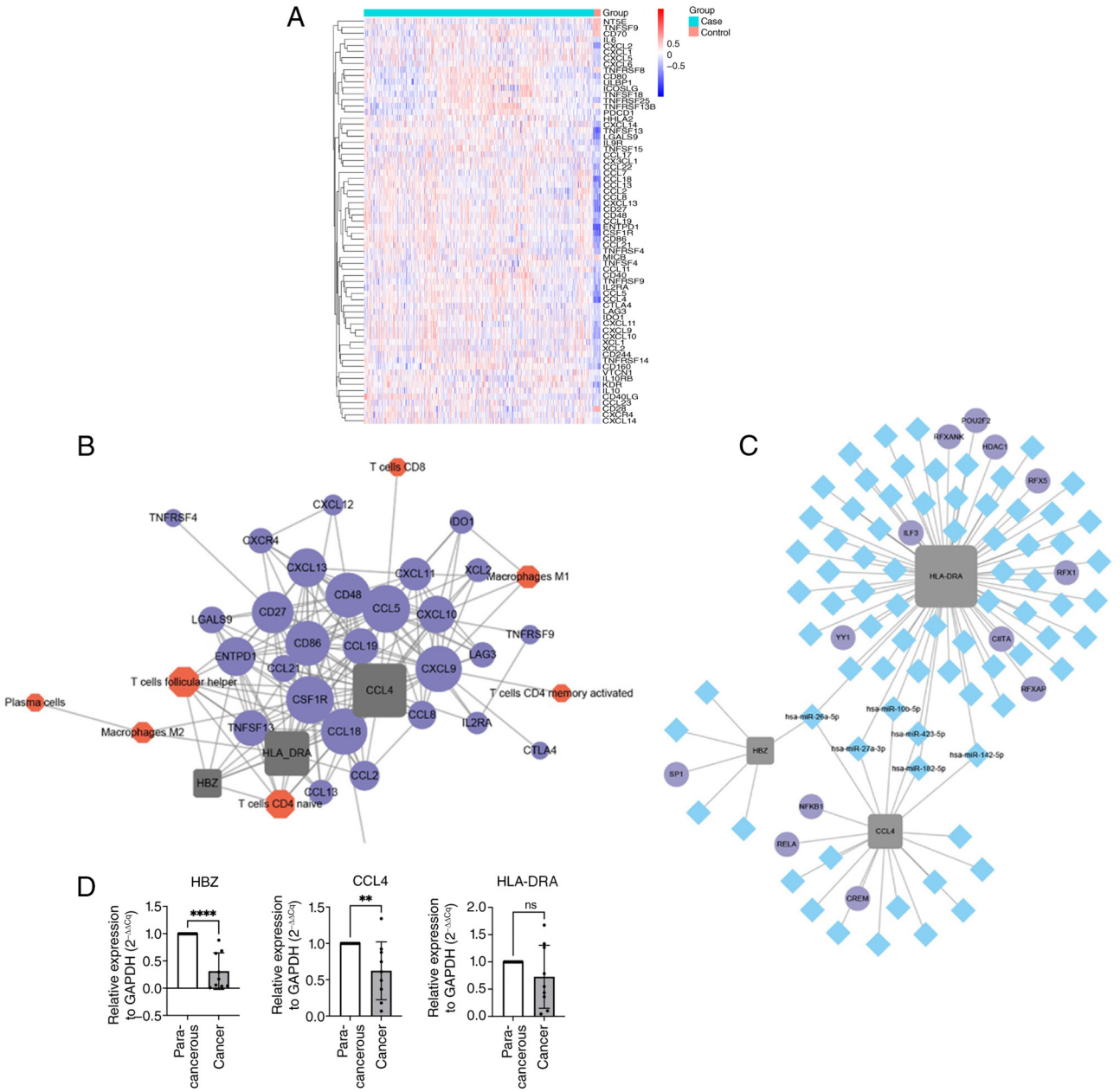

Figure 7. Regulatory network of feature genes. (A) Heatmap showing the differential expression levels of immune factors between NSCLC and normal tissues. The colors above represent sample information and the colors below represent the expression levels of immune factors. Blue indicates low expression and red indicates high expression. (B) Immune network of feature genes, where purple represents differential immune factors, red represents differential immune cells, grey represents feature genes and edges in the network represent correlations >0.5. (C) miRNA-TF regulatory network for feature genes. Grey nodes represent feature genes, purple nodes represent TFs, blue nodes represent miRNAs and edges represent regulatory relationships. (D) Reverse transcription-quantitative PCR detection of feature gene expression in NSCLC and adjacent para-cancerous tissues (n=10 paired samples). HBZ was significantly downregulated (****P<0.0001), CCL4 was significantly downregulated (**P<0.01) and HLA-DRA showed a non-significant downward trend. Expression levels were normalized to GAPDH using the $2^{-\Delta\Delta Cq}$ method. miRNA, microRNA; TF, transcription factor; NSCLC, non-small cell lung cancer; ns, no significance.

with immune cells. For example, the HBZ gene was found to bind directly to CD4$^+$-naïve T cells.

HLA-DRA, a critical component of MHC class II molecules, exhibits a positive correlation between its expression and CD8$^+$ T-cell infiltration (27-29). HLA-DRA plays a central role in antigen presentation to CD4$^+$ T helper cells, which is essential for initiating adaptive antitumor immune responses. Mechanistically, HLA-DRA may interact with ADP-ribosylation via PARP-dependent chromatin remodeling. PARP1-mediated ADP-ribosylation has been shown to regulate the transcription of MHC class II genes by modulating chromatin accessibility at their promoter regions (30,31). This highlights the pivotal role of antigen presentation efficiency in antitumor immunity. However, the present RT-qPCR validation results revealed that HLA-DRA expression in NSCLC tissues did not show a statistically significant difference, only demonstrating a downward trend, likely due to the small sample size in this study. CCL4, a member of the CC chemoattractant family, exhibits dual regulatory properties. It may enhance antitumor immunity by recruiting cytotoxic T cells through the CCL4-CCR5 axis, but it can also be associated with the recruitment of immunosuppressive cells (32-35).

The CCL4-CCR5 signaling axis is particularly noteworthy because $CCR5^+$ cytotoxic T cells recruited to the TME can exert direct cytolytic activity against tumor cells. However, this axis may also recruit $CCR5^+$ Tregs and myeloid-derived suppressor cells, leading to a context-dependent immunological outcome (36,37). This suggests that the spatiotemporal expression pattern of CCL4 determines its ultimate biological effects. Studies have indicated that CCL4 expression is closely associated with the infiltration of $CD8^+$ T cells, as well as with mutations in genes related to lung cancer treatment (e.g., EGFR, ALK and ROS1) (38). These findings suggest that CCL4 may play a role in immune evasion, impacting the efficacy of immunotherapy and targeted therapy, and could potentially serve as a therapeutic target.

Elevated HBZ expression was found to correlate with increased Treg infiltration, suggesting its indirect role in fostering an immunosuppressive microenvironment by modulating cell cycle-related factors. Studies have demonstrated that aberrant HBZ expression is closely linked to tumor immune evasion, suppressing cytotoxic T-cell activity and reinforcing an immunosuppressive environment (39,40). Furthermore, HBZ may enhance tumor cell resistance to immune checkpoint inhibitors by regulating PD-L1 expression, further impairing the host's antitumor immune response (41). In the context of ADP-ribosylation, HBZ may modulate immune cell function by influencing PARP-mediated signaling pathways that control T-cell exhaustion and Treg differentiation (42), thereby contributing to the immunosuppressive microenvironment observed in NSCLC. In summary, the elevated expression of HBZ in NSCLC, together with its capacity to enhance Treg infiltration, modulate PD-L1-mediated immune checkpoint signaling and engage PARP-mediated regulation of T-cell function, supports its role as a candidate ADP-ribosylation-associated biomarker linked to immune evasion in the tumor microenvironment.

This study systematically integrated ADP-ribosylation modification with the regulation of the TME in NSCLC, unraveling the cascade regulatory network of 'ADP-ribosylation-signature genes-immune cells'. This framework provides new insights into the epigenetic-immune crosstalk mechanisms underlying tumor progression. By combining transcriptomic analysis with ELISA protein-level validation and immune infiltration profiling, this integrative approach represents a comprehensive effort to bridge the gap between epigenetic modifications and immune regulation in NSCLC. From a clinical perspective, the current findings carry multifaceted significance. The significantly reduced expression of HBZ in NSCLC suggests its potential as a novel prognostic biomarker. Furthermore, independent ROC analysis in the TCGA LUAD cohort confirmed the discriminative utility of the three-gene signature, with the combined PC1 signature achieving an AUC of 0.772, providing quantitative support for the diagnostic relevance of the machine learning-assisted feature selection strategy employed in this study. Furthermore, targeting the CCL4/CCR5 axis or enhancing HLA-DRA-mediated antigen presentation may offer promising therapeutic strategies for patients who are resistant to immunotherapy.

However, this study has several limitations. The bioinformatics analysis relies on public databases and sample heterogeneity may affect the results. In particular, subtype heterogeneity (LUAD vs. LUSC), disease stage, smoking status and treatment history represent potential confounders that were not systematically adjusted for in this analysis. Additionally, no subtype-specific or stage-specific stratification was performed. To strengthen the findings, future studies should involve multi-cohort validation with explicit batch-effect correction. While the expression patterns of the signature genes were validated through bioinformatics and RT-qPCR, direct mechanistic evidence remains lacking. Furthermore, the clinical RT-qPCR validation cohort was confined to a single institution and comprised only 10 paired samples, which may have introduced selection bias and limited statistical power; In addition, the use of bulk transcriptomic data precludes resolution of cell-type-specific expression patterns, which is particularly relevant for HLA-DRA, whose expression is largely confined to antigen-presenting cells within the tumor microenvironment. Future research should include functional experiments to investigate the effects of these genes on NSCLC cell proliferation and migration, single-cell-resolution analyses to characterize the expression dynamics of HLA-DRA, CCL4 and HBZ across distinct cell populations, as well as immune co-culture assays to elucidate their roles in regulating immune cell function.

In conclusion, this study systematically identified HBZ, HLA-DRA and CCL4 as genes transcriptionally associated with ADP-ribosylation pathway activity that may participate in immune regulation in NSCLC. The ELISA validation of elevated PARP protein levels in NSCLC tissues provides complementary protein-level support for ADP-ribosylation pathway dysregulation. Among these genes, HBZ and CCL4 were experimentally validated by RT-qPCR in an expanded clinical cohort, while HLA-DRA requires further investigation. These findings lay the foundation for future mechanistic and clinical validation studies to uncover the functional networks through which these genes modulate the TME via multiple mechanisms.

Acknowledgements

Not applicable.

Funding

This work was supported by the Clinical Medicine Excellent Talent Program funded by the Hebei Provincial Government (grant no. 20230383).

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

JH and QZ conceived and designed the study, retrieved and curated the public transcriptomic datasets, performed the bioinformatics analyses (including differential expression analysis, weighted gene co-expression network analysis and machine learning-based feature selection), interpreted the integrated bioinformatics and experimental results, and drafted the original manuscript. SL

conceived and designed the study, contributed substantially to the interpretation of the clinical and bioinformatics findings, supervised the experimental work and provided critical revision of the manuscript for important intellectual content. HoZ contributed to the acquisition and curation of clinical specimens and the corresponding clinico-pathological data, and participated in the interpretation of the validation results. HuZ performed the formal analysis of the immune-infiltration data and the single-gene GSEA datasets, and was responsible for data visualization. ZH performed the ELISA experiments, acquired the corresponding protein-level data and contributed to their interpretation. JY and LK performed the RT-qPCR validation experiments, including RNA extraction, reverse transcription and quantitative PCR, and contributed to the interpretation of the experimental results. ZW performed bioinformatics data acquisition and pre-processing of the GEO datasets and assisted with primer design for the RT-qPCR experiments. SY contributed to the construction and interpretation of the miRNA-TF regulatory network and the immune-infiltration analyses. JH and SL confirm the authenticity of all the raw data. All authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Hebei General Hospital (grant no. 2025-LW-0229). All procedures involving human participants were conducted in accordance with the ethical standards of the national research committee and adhered to the 1964 Helsinki Declaration and its subsequent amendments or comparable ethical standards. Written informed consent was obtained from all patients prior to sample collection, including explicit consent for the use of their specimens and data in subsequent research projects.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74: 229-263, 2024.
2. Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W and He J: Cancer incidence and mortality in China, 2022. J Natl Cancer Cent 4: 47-53, 2024.
3. Zeng H, Zheng R, Sun K, Zhou M, Wang S, Li L, Chen R, Han B, Liu M, Zhou J, *et al*: Cancer survival statistics in China 2019-2021: A multicenter, population-based study. J Natl Cancer Cent 4: 203-213, 2024.
4. Relli V, Trerotola M, Guerra E and Alberti S: Abandoning the notion of non-small cell lung cancer. Trends Mol Med 25: 585-594, 2019.
5. Jiang T, Shi J, Dong Z, Hou L, Zhao C, Li X, Mao B, Zhu W, Guo X, Zhang H, *et al*: Genomic landscape and its correlations with tumor mutational burden, PD-L1 expression, and immune cells infiltration in Chinese lung squamous cell carcinoma. J Hematol Oncol 12: 75, 2019.
6. Stewart PA, Welsh EA, Slebos RJC, Fang B, Izumi V, Chambers M, Zhang G, Cen L, Pettersson F, Zhang Y, *et al*: Proteogenomic landscape of squamous cell lung cancer. Nat Commun 10: 3578, 2019.
7. Fois SS, Paliogiannis P, Zinellu A, Fois AG, Cossu A and Palmieri G: Molecular epidemiology of the main druggable genetic alterations in non-small cell lung cancer. Int J Mol Sci 22: 612, 2021.
8. Tsuboi M, Herbst RS, John T, Kato T, Majem M, Grohé C, Wang J, Goldman JW, Lu S, Su WC, *et al*: Overall survival with osimertinib in resected *EGFR*-mutated NSCLC. N Engl J Med 389: 137-147, 2023.
9. Schuller M and Ahel I: Beyond protein modification: The rise of non-canonical ADP-ribosylation. Biochem J 479: 463-477, 2022.
10. Lüscher B, Bütepage M, Eckei L, Krieg S, Verheugd P and Shilton BH: ADP-ribosylation, a multifaceted posttranslational modification involved in the control of cell physiology in health and disease. Chem Rev 118: 1092-1136, 2018.
11. Yang Y, Li X, Xiao S, Wei Q, Ren L, Yao Y and Liu N: PARylation of POLG mediated by PARP1 accelerates ferroptosis-induced vascular calcification via activating Adora2a/Rap1 signaling. Arterioscler Thromb Vasc Biol 45: 1175-1191, 2025.
12. Jiang F, Xu Y, Jiang Z, Hu B, Lv Q and Wang Z: Deciphering the immunological and prognostic features of hepatocellular carcinoma through ADP-ribosylation-related genes analysis and identify potential therapeutic target ARFIP2. Cell Signal 117: 111073, 2024.
13. Aliazis K, Christofides A, Shah R, Yeo YY, Jiang S, Charest A and Boussiotis VA: The tumor microenvironment's role in the response to immune checkpoint blockade. Nat Cancer 6: 924-937, 2025.
14. Tufail M, Jiang CH and Li N: Immune evasion in cancer: Mechanisms and cutting-edge therapeutic approaches. Signal Transduct Target Ther 10: 227, 2025.
15. Chuang CH, Zhen YY, Ma JY, Lee TH, Hung HY, Wu CC, Wang PH, Huang CT, Huang MS, Hsiao M, *et al*: CD47-mediated immune evasion in early-stage lung cancer progression. Biochem Biophys Res Commun 720: 150066, 2024.
16. Robert C, Ribas A, Schachter J, Arance A, Grob JJ, Mortier L, Daud A, Carlino MS, McNeil CM, Lotem M, *et al*: Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): Post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. Lancet Oncol 20: 1239-1251, 2019.
17. Ribeiro VC, Russo LC and Hoch NC: PARP14 is regulated by the PARP9/DTX3L complex and promotes interferon γ-induced ADP-ribosylation. Embo J 43: 2908-2928, 2024.
18. Dasovich M and Leung AKL: PARPs and ADP-ribosylation: Deciphering the complexity with molecular tools. Mol Cell 83: 1552-1572, 2023.
19. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP and Tamayo P: The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 1: 417-425, 2015.
20. Langfelder P and Horvath S: WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9: 559, 2008.
21. Sticht C, De La Torre C, Parveen A and Gretz N: miRWalk: An online resource for prediction of microRNA binding sites. PLoS One 13: e0206239, 2018.
22. Chen Y and Wang X: miRDB: An online database for prediction of functional microRNA targets. Nucleic Acids Res 48: D127-D131, 2020.
23. Agarwal V, Bell GW, Nam JW and Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. Elife 4: e05005, 2015.
24. Feng C, Song C, Liu Y, Qian F, Gao Y, Ning Z, Wang Q, Jiang Y, Li Y, Li M, *et al*: KnockTF: A comprehensive human gene expression profile database with knockdown/knockout of transcription factors. Nucleic Acids Res 48: D93-D100, 2020.
25. Chang L and Xia J: MicroRNA regulatory network analysis using miRNet 2.0. Methods Mol Biol 2594: 185-204, 2023.
26. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods 25: 402-408, 2001.
27. Chang JQ, Guo Y, Yuan WJ, Chen YM, Liu BW, Li WT, Ding XM, Fu XD, Lou YH, Chen ZR, *et al*: HLA-DR$^+$ tumor cells show an association with a distinct immune microenvironment and CD8$^+$ T-cell exhaustion in HBV-associated hepatocellular carcinoma. Adv Sci (Weinh) 12: e02979, 2025.

28. Zhao Y, Chen K, Yang H, Zhang F, Ding L, Liu Y, Zhang L, Zhang Y, Wang H and Deng Y: HLA-DR genetic polymorphisms and hepatitis B virus mutations affect the risk of hepatocellular carcinoma in Han Chinese population. Virol J 20: 283, 2023.
29. Yan C, Ma Y, Li J, Chen X and Ma J: Identification of key immune cell-related genes involved in tumorigenesis and prognosis of cervical squamous cell carcinoma. Hum Vaccin Immunother 19: 2254239, 2023.
30. Ke Y, Zhang J, Lv X, Zeng X and Ba X: Novel insights into PARPs in gene expression: Regulation of RNA metabolism. Cell Mol Life Sci 76: 3283-3299, 2019.
31. Ciccarone F, Zampieri M and Caiafa P: PARP1 orchestrates epigenetic events setting up chromatin domains. Semin Cell Dev Biol 63: 123-134, 2017.
32. Ozga AJ, Chow MT and Luster AD: Chemokines and the immune response to cancer. Immunity 54: 859-874, 2021.
33. Yang C, Geng H, Yang X, Ji S, Liu Z, Feng H, Li Q, Zhang T, Zhang S, Ma X, *et al*: Targeting the immune privilege of tumor-initiating cells to enhance cancer immunotherapy. Cancer Cell 42: 2064-2081.e19, 2024.
34. Tian Y, Zhang L, Ping Y, Zhang Z, Yao C, Shen C, Li F, Wen C and Zhang Y: CCR5 and IL-12 co-expression in CAR T cells improves antitumor efficacy by reprogramming tumor microenvironment in solid tumors. Cancer Immunol Immunother 74: 55, 2025.
35. Shang S, Yang C, Chen F, Xiang RS, Zhang H, Dai SY, Liu J, Lv XX, Zhang C, Liu XT, *et al*: ID1 expressing macrophages support cancer cell stemness and limit $CD8^+$ T cell infiltration in colorectal cancer. Nat Commun 14: 7661, 2023.
36. Schlecker E, Stojanovic A, Eisen C, Quack C, Falk CS, Umansky V and Cerwenka A: Tumor-infiltrating monocytic myeloid-derived suppressor cells mediate CCR5-dependent recruitment of regulatory T cells favoring tumor growth. J Immunol 189: 5602-5611, 2012.
37. Blattner C, Fleming V, Weber R, Himmelhan B, Altevogt P, Gebhardt C, Schulze TJ, Razon H, Hawila E, Wildbaum G, *et al*: $CCR5^+$ myeloid-derived suppressor cells are enriched and activated in melanoma lesions. Cancer Res 78: 157-167, 2018.
38. Chen R, Ma L, Jiang C and Zhang S: Expression and potential role of CCL4 in CD8+T cells in NSCLC. Clin Transl Oncol 24: 2420-2431, 2022.
39. Arnold J, Zimmerman B, Li M, Lairmore MD and Green PL: Human T-cell leukemia virus type-1 antisense-encoded gene, Hbz, promotes T-lymphocyte proliferation. Blood 112: 3788-3797, 2008.
40. Higuchi Y, Yasunaga JI, Mitagami Y, Tsukamoto H, Nakashima K, Ohshima K and Matsuoka M: HTLV-1 induces T cell malignancy and inflammation by viral antisense factor-mediated modulation of the cytokine signaling. Proc Natl Acad Sci USA 117: 13740-13749, 2020.
41. Zhao T: The role of HBZ in HTLV-1-induced oncogenesis. Viruses 8: 34, 2016.
42. Satou Y, Yasunaga J, Zhao T, Yoshida M, Miyazato P, Takai K, Shimizu K, Ohshima K, Green PL, Ohkura N, *et al*: HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo. PLoS Pathog 7: e1001274, 2011.