Cancer Biome and Targeted Therapy
ISSN: 3070-9989
www.cancerbiometherapy.com

RESEARCH ARTICLE · OPEN ACCESS

Integrative Network Toxicology and Multi-Omics Analysis Identifies Candidate Molecular Pathways Linking Acrylonitrile Exposure to Lung Cancer

Wenjing Li[1*], Yaqi Hu[1*], Chaofan Zhang[1], Tiansong Cao[1], Louis Zizhao Wang[3,4], Yanqing Liu[5,6], Hongzhi Yu[7], Yang Cao[2#], Jinrui Dong[1#]

Received: April 30, 2026
Accepted: August 4, 2026
Published: November 15, 2026
DOI: 10.66505/cbtt.v1i4.58

Correspondence: Jinrui Dong (jinrui_dong@tju.edu.cn); Yang Cao (caoyang0801@jlu.edu.cn)
* Equal contributors to this work.

© 2026 The Author(s). Published by GCINC Press, Spokane, Washington, United States. Open Access licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited. To view a copy: https://creativecommons.org/licenses/by/4.0/

Abstract

Background: Acrylonitrile (AN) is an industrial chemical and environmental pollutant with recognized carcinogenic potential. However, the molecular pathways potentially linking AN exposure to lung cancer remain incompletely characterized.

Methods: We integrated network toxicology, machine learning, SHapley Additive exPlanations (SHAP), molecular docking and dynamics simulations, and transcriptomic–metabolomic profiling to identify candidate pathways associated with AN exposure and lung cancer.

Results: The intersection of predicted AN-associated targets, lung cancer-related genes, and tumor-associated differentially expressed genes identified 103 candidate genes. Enrichment analyses highlighted inflammatory and chemokine signaling, NF-κB, MAPK, and calcium signaling. Complementary machine-learning approaches prioritized EZH2, PPARG, PARP1, CXCR2, and EGFR. For tumor-versus-normal classification, the five-gene logistic regression model yielded descriptive areas under the receiver operating characteristic curve of 0.925 in the training dataset and 0.846 in the evaluation dataset. Molecular simulations generated hypotheses concerning potential AN interactions with EZH2 and EGFR. Separately, transcriptomic and metabolomic profiling of AN-exposed A549 cells identified alterations in xenobiotic metabolism, inflammatory signaling, and carbohydrate and lipid metabolism.

Conclusions: These complementary analyses identify candidate molecular pathways at the intersection of AN toxicology and lung cancer biology. Because the computational analyses primarily used existing lung cancer datasets and experimental profiling was restricted to a single cancer cell line, the findings do not establish AN-specific causal mechanisms or clinically validated biomarkers. Exposure-controlled studies in additional pulmonary models and human populations are needed to evaluate the proposed pathways and their translational relevance.

Keywords: Acrylonitrile; lung cancer; environmental toxicology; network toxicology; machine learning; transcriptomics; metabolomics; molecular docking.

1. Introduction

Acrylonitrile (AN) is a high-production-volume industrial chemical extensively used in the manufacture of synthetic fibers, plastics, rubber, and multiple chemical intermediates. Following absorption, AN undergoes biotransformation primarily through glutathione conjugation and oxidative metabolism (1, 2). The oxidative pathway generates the highly reactive epoxide intermediate 2-cyanoethylene oxide, which can covalently interact with proteins and nucleic acids or undergo further conversion to secondary metabolites. These bioactivation events are mechanistically implicated in AN-induced cytotoxicity and tissue injury (1, 2). Current evidence indicates that AN toxicity is characterized by oxidative stress, oxidative DNA damage, and genotoxicity, with recent studies further emphasizing its carcinogenic potential (3, 4). In addition to these molecular toxicities, AN has been associated with multi-organ injury, particularly inhalation-related pulmonary damage. Occupational studies have characterized worker exposure to AN (5) and evaluated lung cancer risk in exposed populations (6, 7). Importantly, biomonitoring studies have detected AN exposure not only in occupational settings but also in the general population through tobacco smoke and other environmental sources (8).

The convergence of epidemiological and mechanistic evidence has heightened concern regarding the pulmonary toxicity and carcinogenicity of AN. Occupational cohorts, systematic reviews, and biomarker studies highlight its occupational and environmental relevance (5–8). Experimental studies demonstrate that AN and its reactive metabolites induce oxidative stress, DNA lesions, and genotoxicity, providing plausible mechanisms for malignant transformation (1–4). However, the molecular mechanisms linking AN exposure to lung carcinogenesis remain poorly defined. Previous investigations have primarily addressed toxicokinetics, oxidative injury, genotoxicity, and epidemiology rather than the coordinated molecular and signaling networks underlying AN-associated lung carcinogenesis (9). Although network-level approaches have advanced research on other environmental carcinogens, systematic network toxicology studies of AN remain limited, particularly those integrating multi-omics analyses with structural evaluation of AN–target interactions (4, 10–12). Consequently, an integrated systems-level framework linking AN exposure to molecular alterations associated with lung carcinogenesis remains lacking.

To address these gaps, we established an integrative framework combining network toxicology, machine learning and SHAP-based interpretation, immune signature analysis, transcriptomic-metabolomic profiling, pathway crosstalk analysis, and molecular dynamics simulation. We hypothesized that AN exposure is associated with coordinated molecular alterations involving inflammatory signaling, oncogenic pathway activation, immune remodeling, DNA damage responses, epigenetic regulation, and metabolic reprogramming. This strategy was designed to reconstruct the molecular response landscape associated with AN exposure, identify key toxicological hub genes and regulatory networks, and prioritize candidate biomarkers of exposure and biological effect. By integrating complementary computational and multi-omics approaches, this study provides a systems-

level framework to elucidate molecular mechanisms potentially linking AN exposure to lung carcinogenesis and to inform future mechanistic validation and environmental health risk assessment.

2. Results

Network Toxicology Analysis of Candidate Genes Linking AN Exposure to Lung Cancer

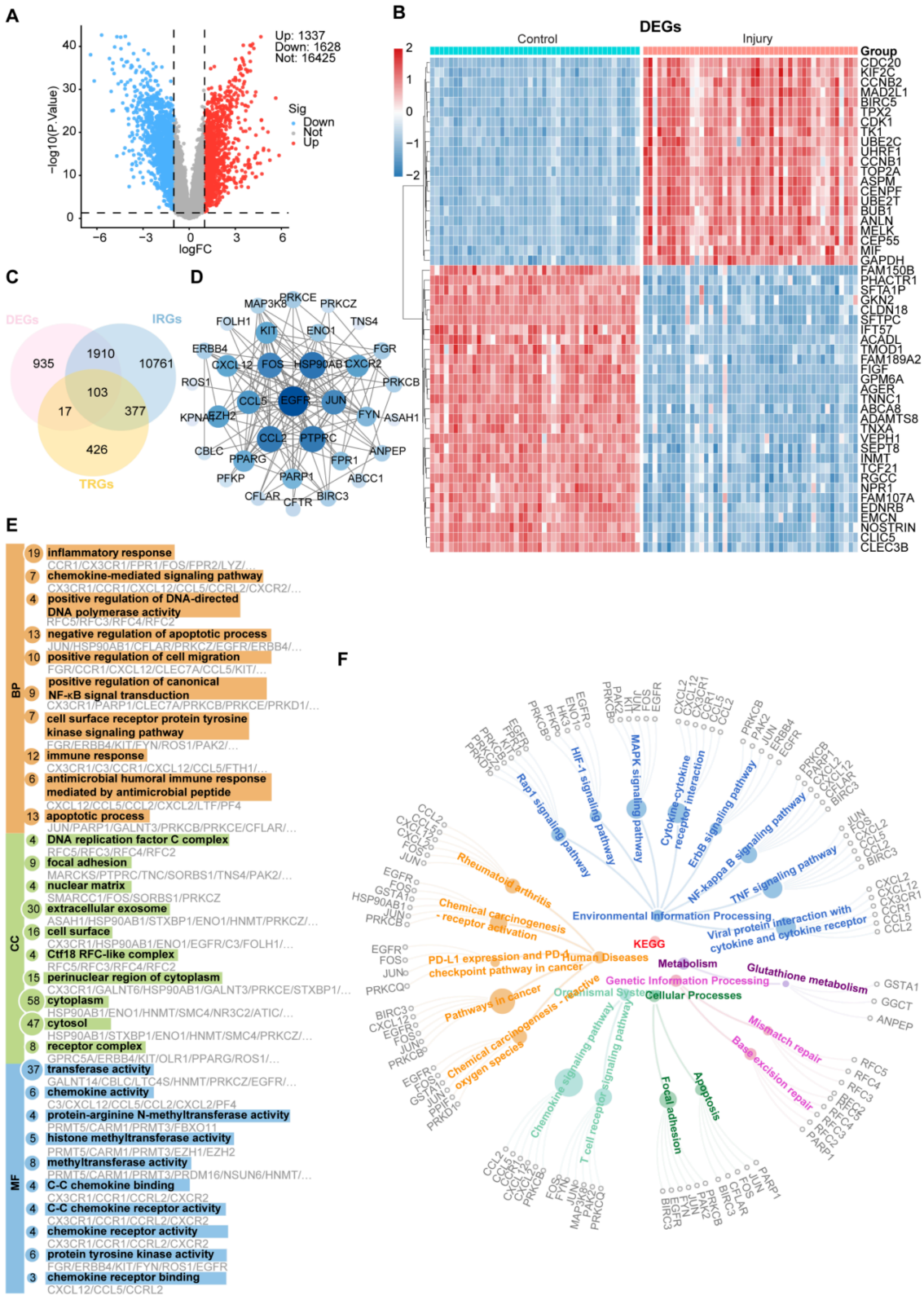


Figure 1. Integrated network toxicology analysis of candidate genes associated with acrylonitrile exposure and lung cancer. *(A) Volcano plot showing differentially expressed genes (DEGs) between tumor and normal samples. Red and blue dots represent upregulated and downregulated genes, respectively, whereas gray dots indicate genes that did not meet the differential expression criteria. (B) Hierarchical clustering heatmap illustrating the expression patterns of representative DEGs across tumor and normal groups. Red and blue indicate relatively high and low expression levels, respectively. (C) Venn diagram showing the intersection of DEGs, inflammation-related genes (IRGs), and toxicology-related genes (TRGs), identifying 103 shared candidate genes. (D) Protein–protein interaction (PPI) network of the intersecting genes, constructed using the STRING database and visualized in Cytoscape. Node size reflects network connectivity, with EGFR identified as a prominent hub. (E) Gene Ontology (GO) enrichment analysis showing representative biological process (BP), cellular component (CC), and molecular function (MF) terms associated with the candidate genes. (F) Circular network visualization of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment, illustrating associations between candidate genes and enriched signaling pathways, including inflammation-related, cancer-associated, and DNA damage response pathways.* ***Abbreviations:*** *AN, acrylonitrile; BP, biological process; CC, cellular component; DEG, differentially expressed gene; GO, Gene Ontology; IRG, inflammation-related gene; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, molecular function; PPI, protein–protein interaction; TRG, toxicology-related gene.*

Transcriptomic analysis identified 2,965 DEGs between tumor and normal samples, including 1,337 upregulated and 1,628 downregulated genes (Figure 1A), with clear separation between groups in hierarchical clustering (Figure 1B). By integrating these DEGs with 923 AN-related target genes and 13,151 lung cancer-related genes, 103 overlapping interGenes were identified (Figure 1C). STRING-based PPI analysis revealed several central nodes, including PTPRC, EGFR, FN1, JUN, FOS, CXCL8, and CXCL12 (Figure 1D). Functional enrichment analysis showed that these interGenes were mainly associated with inflammatory response, chemokine-mediated signaling, cell migration, apoptosis regulation, and NF-κB signaling. GO terms were enriched in the extracellular exosome, cell surface, focal adhesion, and cytosol, with molecular functions related to chemokine activity, receptor-ligand activity, chemokine receptor binding, and protein tyrosine kinase activity (Figure 1E). KEGG analysis identified enrichment of cytokine–cytokine receptor interactions, TNF, NF-κB, chemokine, IL-17, MAPK, calcium signaling, and focal adhesion pathways (Figure 1F). These associations identify biological processes shared by the candidate gene set and lung cancer but do not establish their activation by AN exposure. The 103 intersecting genes were derived partly from lung cancer datasets. Enrichment therefore cannot, by itself, demonstrate AN-induced pathway activation.

Machine Learning-Based Prioritization of Candidate Genes Associated with AN and Lung Cancer

To prioritize candidate genes at the intersection of predicted AN-associated targets and lung cancer biology, three complementary machine-learning algorithms, LASSO, SVM-RFE, and random forest, were applied to the candidate gene set. LASSO regression retained nine genes at the minimum binomial deviance (Suppl Figure S1A-B). SVM-RFE was then used to evaluate the discriminative performance of different feature subsets and identified a compact five-gene subset with strong classification performance (Suppl Figure S1C-D). Random forest analysis further ranked candidate genes according to variable importance (Suppl Figure S1E). Based on the combined evidence from LASSO selection, SVM-RFE-based performance evaluation, and random forest importance ranking, five genes, EZH2, PPARG, PARP1, CXCR2, and EGFR, were prioritized as candidate genes potentially linking AN exposure to lung cancer (Suppl Figure S1F). These five genes were subsequently used for model interpretation, logistic regression-based risk score construction, and downstream mechanistic analyses.

SHAP-Based Interpretation and Tumor-versus-Normal Classification Using Candidate Genes

To explore the classification patterns and relative contributions of the five prioritized genes, eight supervised algorithms, DTS, glmBoost, KNN, logistic regression, neural network, PLS, random forest (RF), and SVM, were compared (Figure 2A).

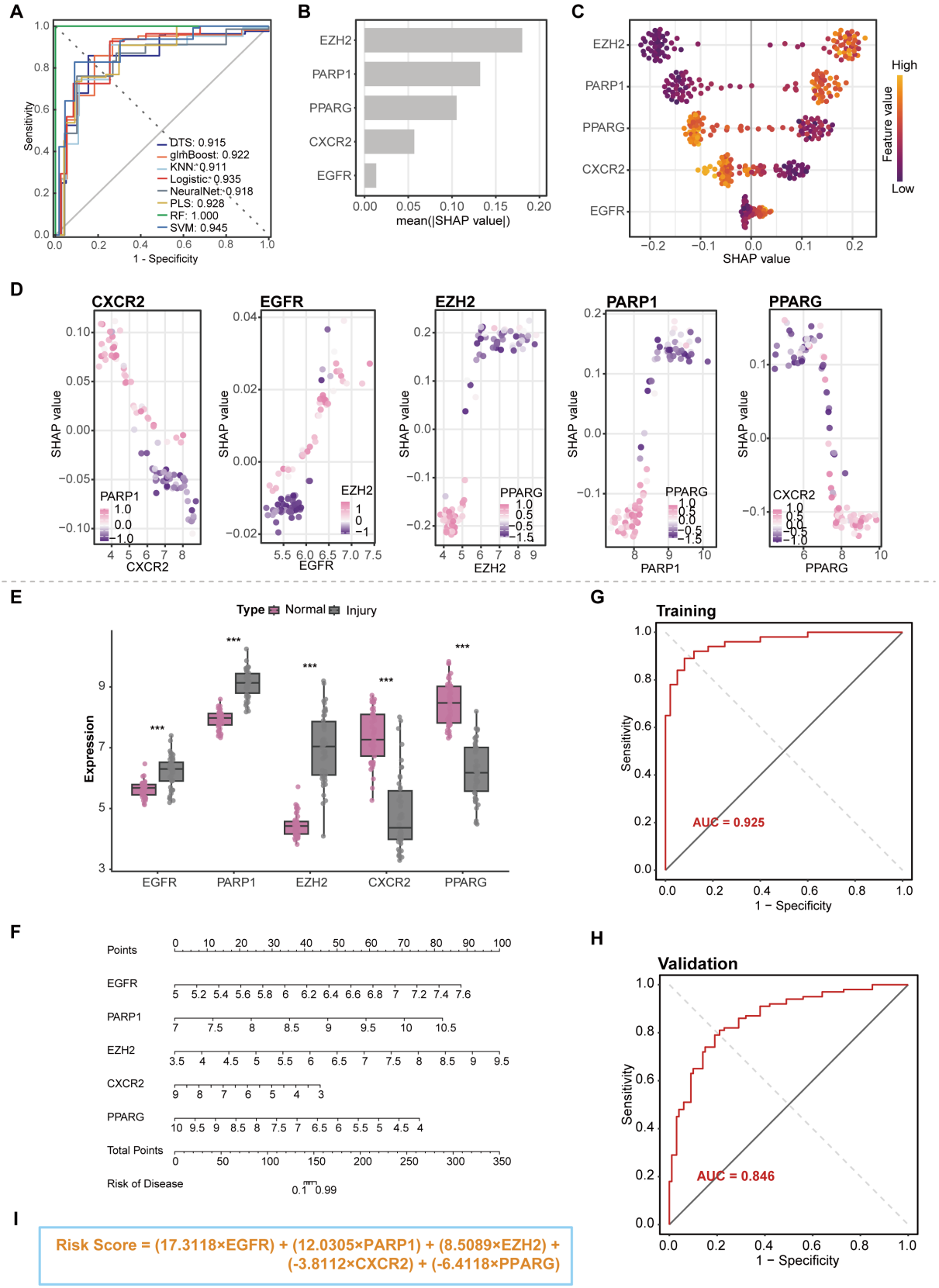


Figure 2. Exploratory SHAP analysis and five-gene tumor-versus-normal classification. *(A) Descriptive comparison of eight supervised machine-learning models using the five prioritized genes. The RF AUC of 1.000 is presented as an exploratory result and does not establish independently validated predictive performance. (B) Mean absolute SHAP values of the five genes in the RF model. (C) SHAP summary plot showing the direction and magnitude of feature contributions. (D) SHAP dependence plots illustrating relationships between gene expression and model predictions. (E) Expression differences*

between normal and tumor samples. (F) Nomogram derived from the five-gene logistic regression model. (G, H) Receiver operating characteristic curves for the five-gene logistic regression model in the training dataset (GSE18842, AUC = 0.925) and the evaluation dataset (GSE32863, AUC = 0.846), respectively. The label "Validation" in panel H refers to the GSE32863 evaluation dataset. Both AUCs are descriptive estimates of tumor-versus-normal discrimination. The independence of feature selection from model evaluation could not be fully verified. The label "Validation" in panel H refers to the GSE32863 evaluation dataset. (I) Five-gene logistic regression score formula.

RF yielded an AUC of 1.000 in this analysis. However, because the original analysis workflow could not be fully verified, this result was considered exploratory and was not interpreted as evidence of independently validated predictive performance. RF was used for subsequent SHAP analysis to characterize the contributions of individual genes to model predictions. The classifier comparison and SHAP analysis were therefore used for exploratory model interpretation rather than to establish the generalizability of the five-gene signature. SHAP analysis identified EZH2 and PARP1 as the main contributors, followed by PPARG, CXCR2, and EGFR (Figure 2B). Consistent with these results, the SHAP summary plot illustrated the distribution and magnitude of SHAP values across the cohort, thereby reflecting both the direction and strength of each gene's contribution to the model output (Figure 2C). Furthermore, SHAP dependence plots delineated the quantitative relationships between gene expression levels and their corresponding contributions to the model predictions (Figure 2D).

The expression patterns of the five selected genes are shown in Figure 2E. PARP1, EZH2, and EGFR were significantly upregulated in tumor samples, whereas CXCR2 and PPARG were downregulated. Based on these patterns, a five-gene logistic regression classifier was developed to distinguish tumor from normal samples, and a nomogram was constructed to visualize individual gene contributions and predicted tumor probabilities (Figure 2F). The five-gene logistic regression model yielded descriptive AUCs of 0.925 in the training dataset and 0.846 in the evaluation dataset (Figures 2G–H). The AUCs summarize tumor-versus-normal discrimination in the two datasets. Independent validation of the complete modeling pipeline requires documented separation of feature selection, model fitting, and evaluation. Prognostic utility, AN-exposure specificity, and clinical applicability require separate evaluation. A weighted five-gene score was calculated using the regression coefficients assigned to each gene. EGFR, PARP1, and EZH2 contributed positively to the score, whereas CXCR2 and PPARG contributed negatively, consistent with their respective expression patterns in tumor and normal samples. The resulting formula was: Risk Score = (17.3118 × EGFR) + (12.0305 × PARP1) + (8.5089 × EZH2) − (3.8112 × CXCR2) − (6.4118 × PPARG). Gene-expression values were at the gene-symbol level, quantile-normalized, and on the log2 scale, without additional z-score standardization. The displayed score represents the gene-expression component of the logistic linear predictor, with the intercept omitted, rather than a calibrated probability (Figure 2I). The full model's linear predictor was calculated as −175.4992 + Risk Score.

Molecular Dynamics Simulations of Predicted AN–Protein Complexes

The three-dimensional structure of acrylonitrile (AN) was retrieved from the PubChem database and subjected to geometry optimization using the MMFF94 molecular mechanics force field to obtain an energetically favorable conformation for subsequent molecular docking. The experimentally determined crystal structure of human enhancer of zeste homolog 2 (EZH2) was obtained from the Protein Data Bank (PDB) and prepared for docking to ensure its suitability for evaluating potential ligand–protein interactions. Molecular docking was performed using AutoDock Vina to investigate the potential binding mode, predicted binding affinity, and spatial orientation of AN within the selected EZH2 binding region. The resulting docking poses were examined to characterize potential interactions between AN and surrounding amino acid residues, providing a structural basis for assessing the plausibility of AN-EZH2 association. The predicted binding configuration and intermolecular

interaction patterns were subsequently visualized and analyzed using Discovery Studio and PyMOL (Figure 3A).

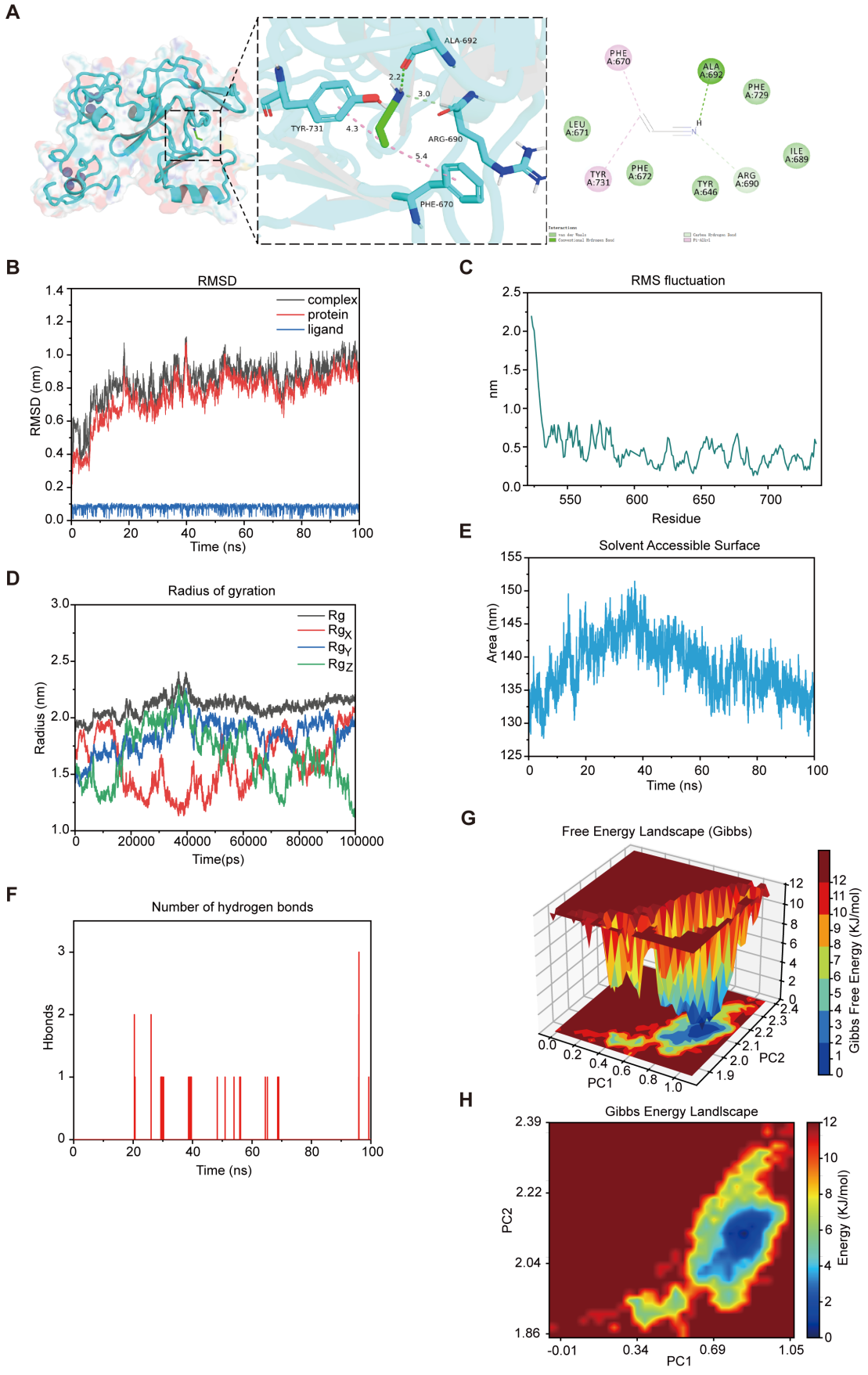


Figure 3. Molecular docking and molecular dynamics analyses of the predicted acrylonitrile–EZH2 complex. *(A) Predicted binding configuration of acrylonitrile (AN) with enhancer of zeste homolog 2 (EZH2), showing the overall three-dimensional binding pose, an enlarged view of the binding region with interacting amino acid residues and predicted intermolecular distances, and a two-dimensional ligand–protein interaction diagram. (B) Root mean square deviation (RMSD) trajectories of the AN–EZH2 complex, protein, and ligand during the 100 ns molecular dynamics (MD) simulation. (C) Root mean square fluctuation (RMSF) profile illustrating residue-specific flexibility of EZH2. (D) Radius of gyration (Rg) and its directional components (Rgx, Rgy, and Rgz), reflecting changes in protein compactness and structural organization throughout the simulation. (E) Solvent-accessible surface area (SASA) profile showing temporal changes in solvent exposure. (F) Time-dependent hydrogen bond profile illustrating the number and occurrence of hydrogen bonds between AN and EZH2 during the simulation. (G) Three-dimensional Gibbs free energy landscape derived from principal component analysis (PCA), illustrating the conformational energy distribution along the first two principal components (PC1 and PC2). (H) Corresponding two-dimensional Gibbs free energy landscape showing the distribution of sampled conformational states, with blue indicating lower-energy regions and red indicating higher-energy regions.* **Abbreviations:** *AN, acrylonitrile; EZH2, enhancer of zeste homolog 2; MD, molecular dynamics; PCA, principal component analysis; Rg, radius of gyration; RMSD, root mean square deviation; RMSF, root mean square fluctuation; SASA, solvent-accessible surface area.*

To further investigate the conformational stability and dynamic behavior of the predicted AN–EZH2 complex under simulated physiological conditions, a 100 ns molecular dynamics (MD) simulation was performed using GROMACS. The initial complex was subjected to energy minimization to eliminate unfavorable steric contacts and reduce structural strain, followed by sequential equilibration under constant particle number, volume, and temperature (NVT) and constant particle number, pressure, and temperature (NPT) ensembles.

Following equilibration, a 100 ns production simulation was conducted to characterize the time-dependent structural behavior of the complex. Root mean square deviation (RMSD) was calculated to assess overall structural stability, while root mean square fluctuation (RMSF) was used to characterize residue-specific flexibility. The radius of gyration (Rg) and solvent-accessible surface area (SASA) were analyzed to evaluate protein compactness and changes in solvent exposure, respectively. Hydrogen bond profiling was performed to examine the persistence and temporal variation of potential intermolecular interactions.

Additionally, principal component analysis (PCA) was employed to characterize dominant collective motions, while Gibbs free energy landscape analysis was used to identify energetically favorable conformational states and assess the conformational distribution sampled during the simulation. Collectively, these complementary analyses provided a framework for evaluating the predicted structural stability, molecular interactions, and dynamic properties of the AN–EZH2 complex (Figures 3B–H).

As a complementary analysis, molecular docking and a 100-ns molecular dynamics simulation were also performed for the AN–EGFR complex (Suppl. Figure S2). RMSD, RMSF, Rg, SASA, hydrogen-bond dynamics, PCA, and Gibbs free-energy landscape analyses were used to characterize the predicted complex and its conformational behavior during the simulation.

GSVA and Immune-Related Signature Analysis in Lung Cancer

GSVA was performed to investigate pathways associated with CXCR2, EGFR, EZH2, PARP1, and PPARG in the analyzed lung cancer transcriptomic dataset. The results highlighted pathways related to inflammation, proliferation, signal transduction, metabolism, ion transport, autophagy, and membrane transport (Suppl. Figure S3). These associations suggest candidate biological processes involving the five prioritized genes but do not establish their functional activity or regulation by AN exposure. ssGSEA was subsequently used to characterize immune-related transcriptional signatures in the lung cancer dataset. Multiple immune-related signatures showed strong correlations, suggesting coordinated

immune-related transcriptional patterns (Figure 4A). Compared with normal samples, tumor samples showed lower scores for several immune-related signatures and a higher chemokine response score (Figure 4B). Spearman analysis showed that CXCR2 and PPARG were generally positively correlated with immune-related signatures, whereas EGFR, PARP1, and EZH2 were predominantly negatively correlated (Figure 4C-D), indicating potential associations between the five core genes and immune-related transcriptional features.

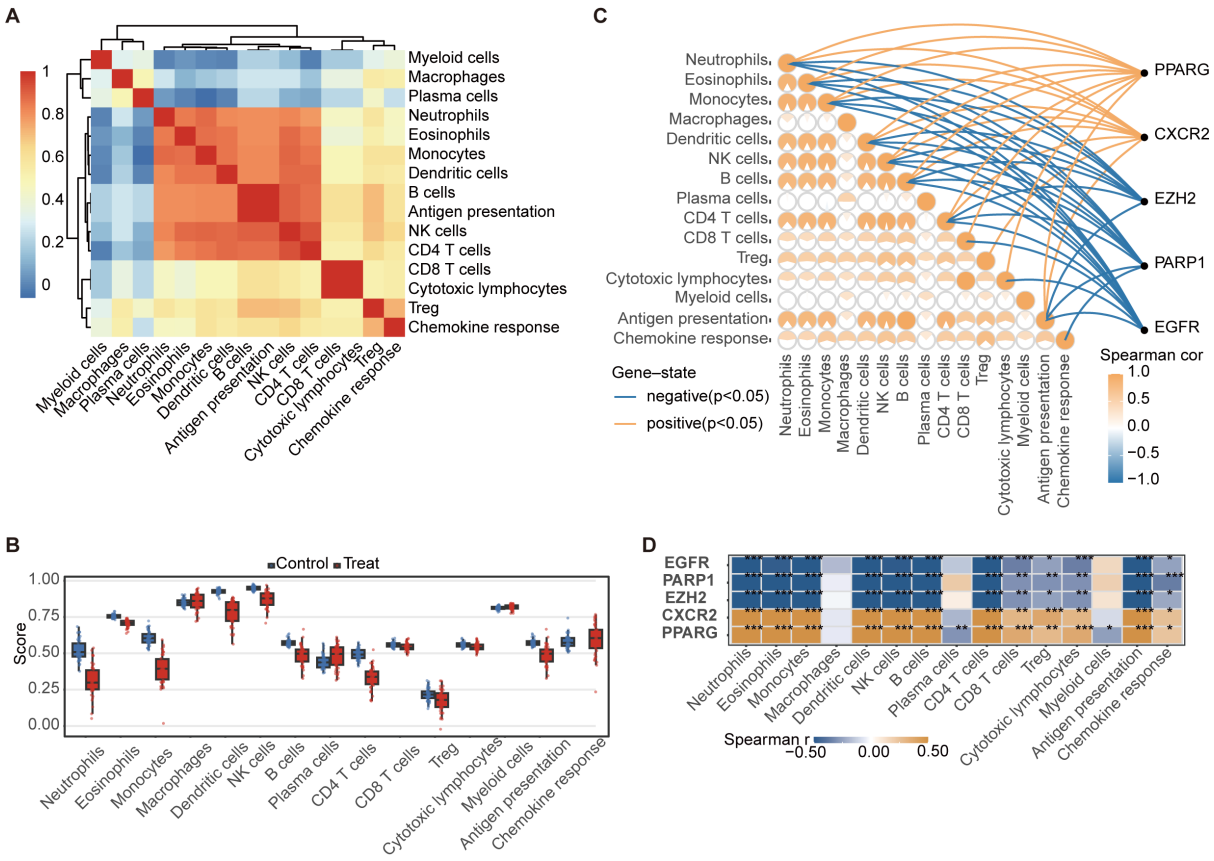


Figure 4. Immune-Related Signature and Correlation Analyses in Lung Cancer. *(A) Spearman correlation heatmap of immune cell and immune-related signatures. (B) Boxplots comparing ssGSEA immune scores between normal and tumor groups. (C) Network-matrix plot showing correlations between the five feature genes and immune signatures; line color indicates correlation direction, and dot intensity indicates correlation strength. (D) Spearman correlation heatmap of five feature genes and immune signatures, with asterisks indicating statistical significance.*

Functional Prediction and Regulatory Network Analysis of Candidate Genes

To further characterize the functional relationships and potential molecular interactions among the five prioritized candidate genes, a GeneMANIA network was constructed using EGFR, EZH2, PARP1, PPARG, and CXCR2 as input genes. The resulting network revealed extensive functional associations among these candidates and additional interacting molecules, including SUZ12, EED, PIK3CA, BRCA1, CXCL8, CXCL3, and EGF. These associations suggest potential connections among growth factor signaling, epigenetic regulation, DNA damage responses, and inflammatory signaling. Functional annotation identified enrichment in several biological processes, including ERBB signaling, epithelial cell proliferation, histone lysine methylation, regulation of DNA damage responses, and receptor metabolic processes. Collectively, these findings highlight the functional diversity and interconnectedness of the five candidate genes and suggest that their potential contributions to AN-associated lung carcinogenesis may involve coordinated alterations in proliferative signaling, chromatin

regulation, genomic maintenance, and inflammatory responses. However, these network-derived associations remain predictive and require experimental validation (Figure 5A).

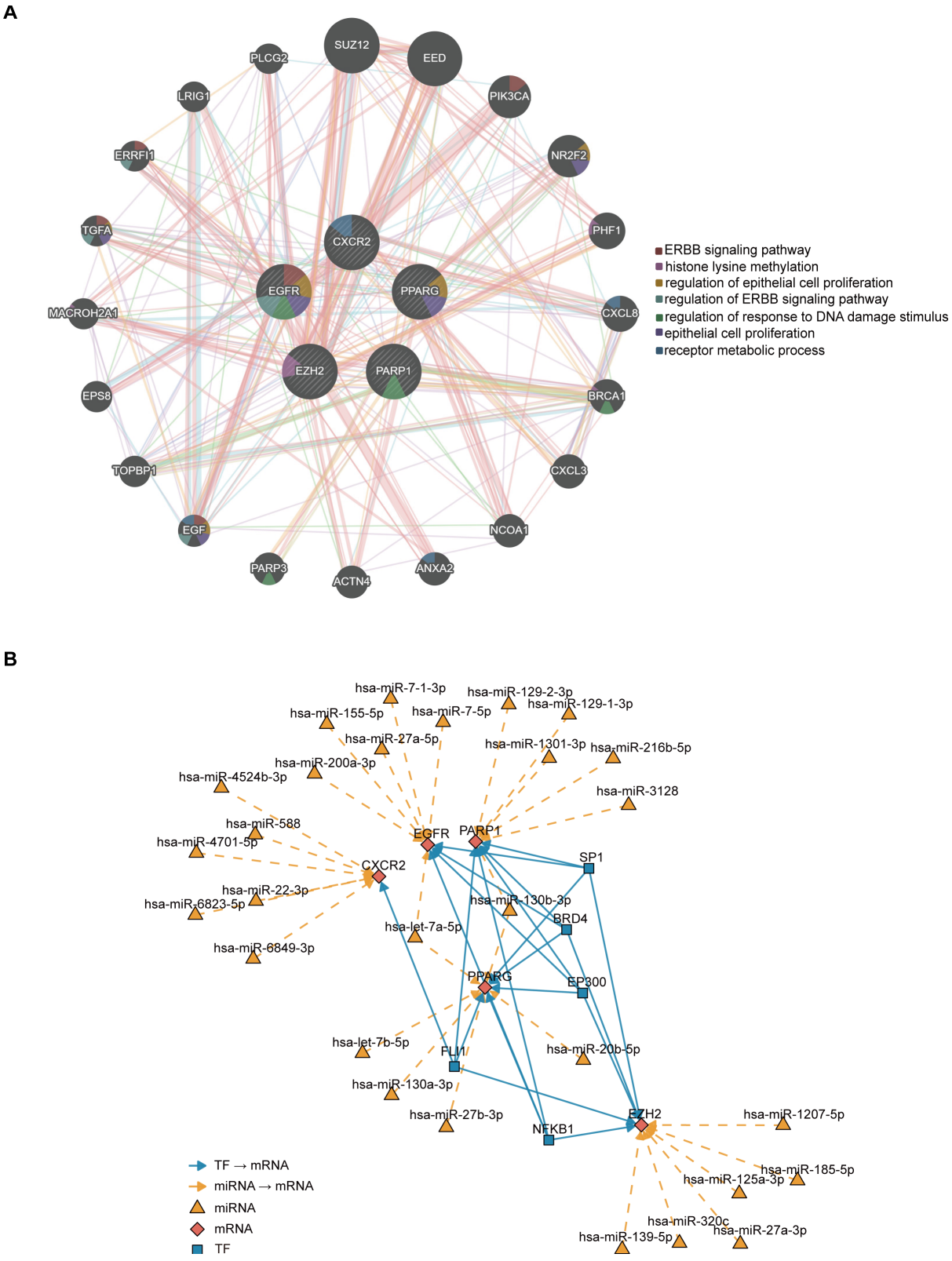


Figure 5. Predicted functional associations and transcriptional and post-transcriptional regulatory networks of the five prioritized candidate genes. *(A) GeneMANIA-derived functional interaction network generated using EGFR, EZH2, PARP1, PPARG, and CXCR2 as input genes. Colored node sectors indicate annotated functions, including ERBB signaling, histone lysine methylation, regulation of epithelial cell proliferation, regulation of ERBB signaling, regulation of the response to DNA*

damage, epithelial cell proliferation, and receptor metabolic processes. Connecting edges represent functional associations among genes. (B) Integrated microRNA (miRNA)–transcription factor (TF)-messenger RNA (mRNA) regulatory network centered on the five prioritized candidate genes, EGFR, EZH2, PARP1, PPARG, and CXCR2. The network illustrates predicted transcriptional and post-transcriptional regulatory interactions involving upstream TFs and miRNAs. Red diamonds represent mRNAs, blue squares indicate TFs, and orange triangles denote miRNAs. Solid blue arrows indicate predicted TF-mRNA regulatory interactions, whereas dashed orange arrows represent predicted miRNA-mRNA interactions. **Abbreviations:** *extracellular signal-regulated kinase; miRNA, microRNA; mRNA, messenger RNA; TF, transcription factor.*

These functional associations highlight potential molecular pathways involving growth factor signaling, epigenetic regulation, epithelial proliferation, inflammatory responses, and DNA damage regulation. To further investigate the potential upstream regulatory mechanisms governing the five prioritized candidate genes, an integrated miRNA–transcription factor (TF)–mRNA network was constructed (Figure 5B). The resulting network revealed extensive predicted regulatory interactions connecting EGFR, EZH2, PARP1, PPARG, and CXCR2 with multiple miRNAs and TFs, suggesting coordinated transcriptional and post-transcriptional regulation. Among the identified regulators, SP1, BRD4, EP300, NFKB1, and FLI1 formed an interconnected TF module, with prominent predicted associations involving PPARG and EZH2.

In parallel, several miRNAs, including hsa-miR-7-1-3p, hsa-miR-155-5p, hsa-miR-129-1-3p, hsa-miR-216b-5p, hsa-miR-130b-3p, hsa-miR-27b-3p, hsa-miR-125a-3p, and hsa-miR-185-5p, were predicted to target the core mRNAs. These regulatory relationships suggest potential mechanisms through which transcriptional and post-transcriptional processes may influence the expression of genes involved in proliferative signaling, chromatin regulation, DNA damage responses, and inflammation. Collectively, the integrated network provides a framework for identifying candidate upstream regulators of the five prioritized genes and their potential contributions to lung carcinogenesis. However, these computationally predicted interactions do not establish direct regulatory activity or AN-exposure specificity, and their biological relevance requires experimental validation.

Transcriptomic and Metabolomic Alterations Induced by AN in A549 cells

Transcriptomic profiling revealed substantial alterations in gene expression in A549 lung adenocarcinoma cells following acrylonitrile (AN) exposure, indicating a broad transcriptional response to AN treatment. Comparative analysis identified 12,646 genes shared between control and AN-treated cells, whereas 527 and 475 genes were uniquely detected in the control and AN-exposed groups, respectively (Figure 6A). These findings suggest that AN exposure induces detectable changes in the transcriptional landscape of A549 cells, potentially reflecting cellular responses to chemical stress and associated molecular perturbations. A549 cells provide a tractable experimental model for investigating AN-responsive molecular alterations and identifying candidate genes and signaling pathways potentially involved in its pulmonary toxicity.

However, because these cells are already malignant, the observed transcriptional changes cannot establish the molecular events through which AN initiates malignant transformation in normal pulmonary epithelial cells. Furthermore, using a single exposure concentration and duration limits the ability to distinguish transient adaptive stress responses from sustained molecular alterations potentially relevant to chronic environmental exposure. Additional studies using nonmalignant pulmonary epithelial cells, multiple exposure concentrations, and longitudinal sampling are therefore necessary to establish the temporal and biological relevance of these transcriptional responses.

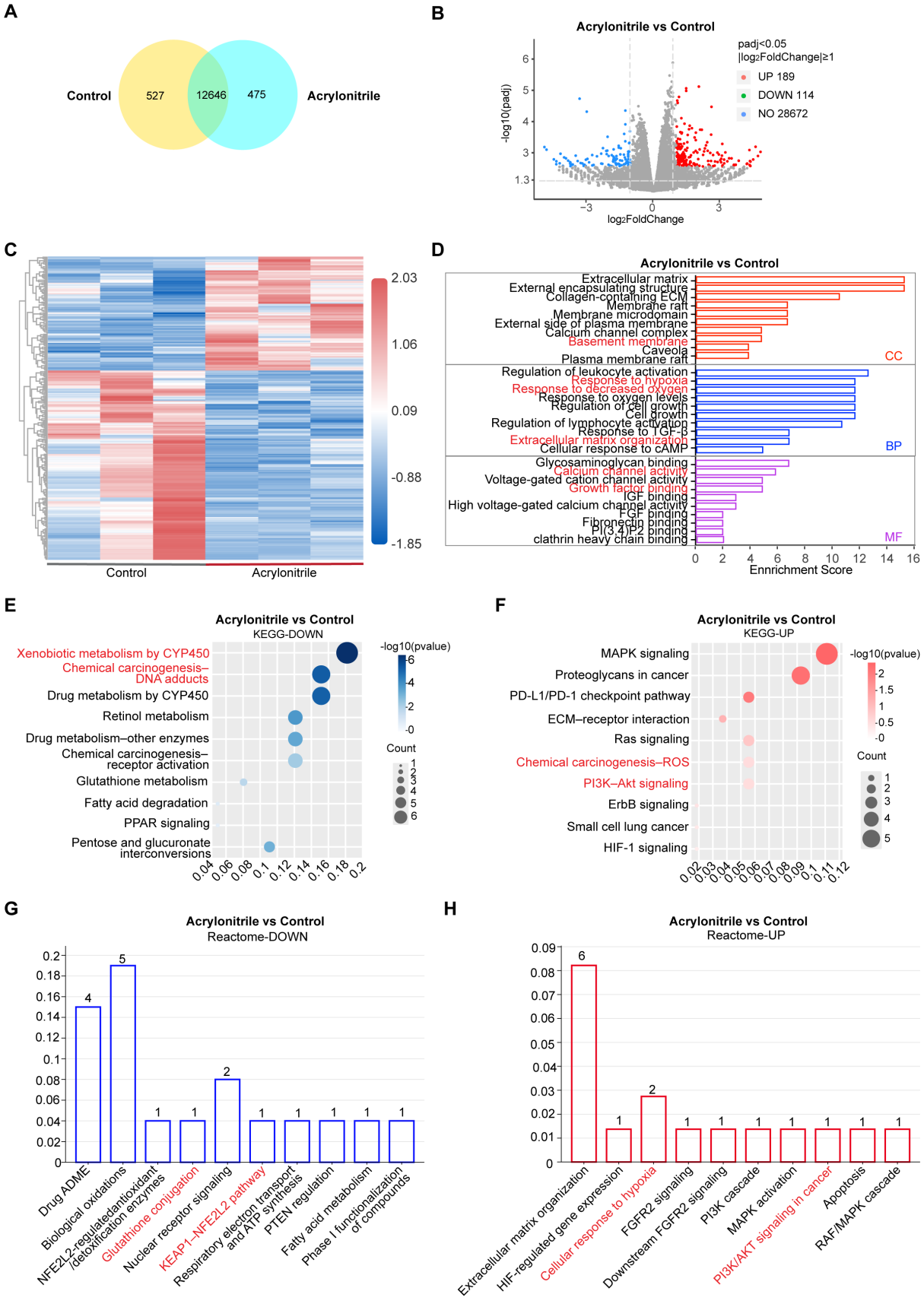


Figure 6. Transcriptomic profiling and functional enrichment analysis of A549 lung adenocarcinoma cells following acrylonitrile exposure. *(A) Venn diagram showing the distribution of detected genes in control and acrylonitrile (AN)-treated cells, including 12,646 shared genes and 527 and 475 genes uniquely detected in the control and AN-treated groups,*

respectively. (B) Volcano plot illustrating differentially expressed genes (DEGs) following AN exposure, with 189 upregulated and 114 downregulated genes identified using an adjusted P-value < 0.05 and an absolute log_2 fold change ≥ 1. Red and green dots represent significantly upregulated and downregulated genes, respectively, while blue dots indicate genes that do not meet the differential expression criteria. (C) Hierarchical clustering heatmap showing DEG expression patterns across control and AN-treated samples. Red and blue indicate relatively high and low expression levels, respectively. (D) Gene Ontology (GO) enrichment analysis of DEGs, showing representative cellular component (CC), biological process (BP), and molecular function (MF) categories, including extracellular matrix organization, responses to oxidative stress, regulation of cell growth, and calcium channel activity. (E) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of downregulated genes, highlighting xenobiotic metabolism by cytochrome P450, chemical carcinogenesis, DNA adduct formation, and glutathione metabolism. (F) KEGG pathway enrichment analysis of upregulated genes, including MAPK signaling, PI3K–Akt signaling, proteoglycans in cancer, and chemical carcinogenesis associated with reactive oxygen species. Bubble size represents gene count, and color indicates enrichment significance. (G) Reactome pathway enrichment analysis of downregulated genes, including biological oxidation, glutathione conjugation, and NRF2-mediated antioxidant responses. (H) Reactome pathway enrichment analysis of upregulated genes, including extracellular matrix organization, cellular responses to hypoxia, and PI3K/AKT signaling in cancer. **Abbreviations:** *AN, acrylonitrile; BP, biological process; CC, cellular component; DEG, differentially expressed gene; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; MF, molecular function; NRF2, nuclear factor erythroid 2-related factor 2; PI3K, phosphoinositide 3-kinase.*

Differential expression analysis identified 189 upregulated and 114 downregulated genes following AN treatment (Figure 6B). Hierarchical clustering revealed distinct expression patterns between control and AN-treated cells, indicating substantial transcriptional alterations in response to AN exposure (Figure 6C). Gene Ontology (GO) enrichment analysis showed that the differentially expressed genes (DEGs) were predominantly associated with extracellular matrix-related components, leukocyte activation, responses to hypoxia and xenobiotics, cytokine regulation, extracellular matrix organization, glycosaminoglycan binding, calcium ion binding, and growth factor binding (Figure 6D). These enriched biological processes and molecular functions suggest that AN exposure may influence extracellular matrix remodeling, cellular stress responses, and growth factor-associated signaling.

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome enrichment analyses further revealed distinct functional associations between upregulated and downregulated genes. Downregulated genes were primarily associated with xenobiotic and drug metabolism, glutathione metabolism, fatty acid degradation, and peroxisome proliferator-activated receptor (PPAR)-related pathways, suggesting potential alterations in cellular detoxification and metabolic regulation. Conversely, upregulated genes were enriched in mitogen-activated protein kinase (MAPK), Ras, PI3K–Akt, ErbB, and hypoxia-inducible factor 1 (HIF-1) signaling, as well as extracellular matrix-related and immune checkpoint-associated pathways (Figures 6E–H).

Collectively, these findings identify candidate AN-responsive molecular processes involving metabolic regulation, oxidative stress responses, proliferative signaling, and extracellular matrix remodeling. However, transcriptomic enrichment reflects associations based on gene expression and does not directly establish changes in pathway activation, metabolic flux, or immune checkpoint function. Additional functional studies are required to determine the biological consequences of these transcriptional alterations.

Metabolomic profiling further revealed metabolic perturbations in AN-treated A549 cells, with 28 upregulated and 21 downregulated metabolites compared with controls (Figure 7A), and clear separation between groups in hierarchical clustering (Figure 7B).

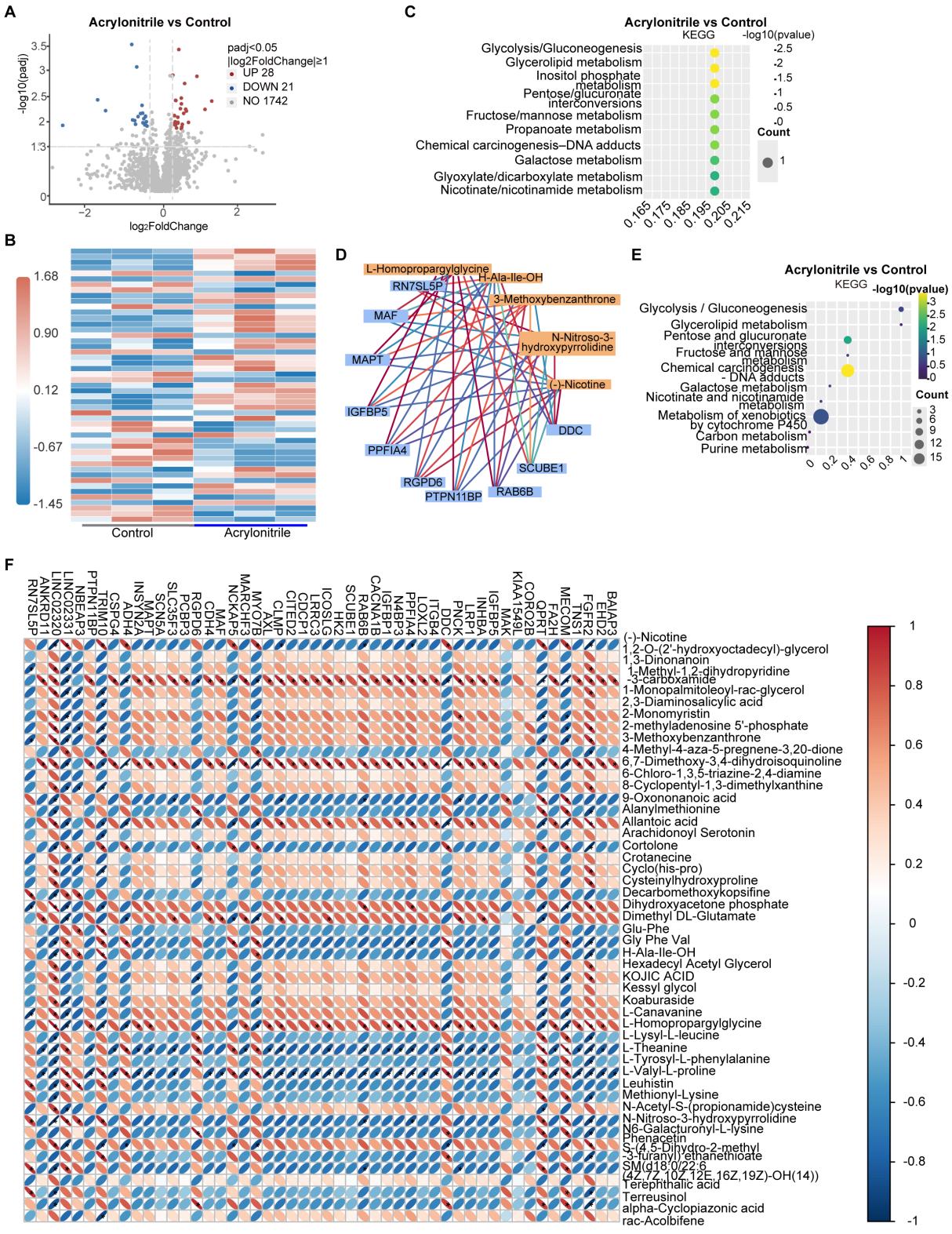


Figure 7. Metabolomic alterations and integrated transcriptomic–metabolomic analysis of A549 lung adenocarcinoma cells following acrylonitrile exposure. *(A) Volcano plot showing differential metabolite abundance between control and acrylonitrile (AN)-treated cells. Red and blue dots represent 28 significantly increased and 21 significantly decreased metabolites, respectively, whereas gray dots indicate metabolites that did not meet the significance criteria (VIP > 1, unadjusted P < 0.05, and absolute log_2 fold change ≥ 1). (B) Hierarchical clustering heatmap illustrating differential metabolite abundance across control and AN-treated samples. Red and blue indicate relatively high and low metabolite abundance, respectively. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of differential metabolites, highlighting glycolysis/gluconeogenesis, glycerolipid metabolism, pentose and glucuronate interconversions,*

chemical carcinogenesis–DNA adducts, and nicotinate/nicotinamide metabolism. (D) Integrated transcriptomic–metabolomic interaction network illustrating associations between differentially expressed genes (DEGs) and altered metabolites. Blue nodes represent genes, orange nodes indicate metabolites, and connecting edges depict their inferred associations. (E) KEGG pathway enrichment analysis of the integrated transcriptomic–metabolomic dataset, highlighting metabolic pathways associated with glycolysis/gluconeogenesis, glycerolipid metabolism, chemical carcinogenesis–DNA adducts, and xenobiotic metabolism by cytochrome P450. In panels C and E, bubble size represents the number of associated features, and color indicates enrichment significance. (F) Correlation matrix illustrating associations between DEGs and differential metabolites. Red and blue indicate positive and negative correlations, respectively, with color intensity reflecting correlation strength.
Abbreviations: *AN, acrylonitrile; DEG, differentially expressed gene; KEGG, Kyoto Encyclopedia of Genes and Genomes.*

KEGG enrichment showed that altered metabolites were mainly involved in carbohydrate and lipid metabolism, including glycolysis/gluconeogenesis, glycerolipid metabolism, inositol phosphate metabolism, pentose and glucuronate interconversions, fructose and mannose metabolism, propanoate metabolism, galactose metabolism, glyoxylate and dicarboxylate metabolism, nicotinate and nicotinamide metabolism, and chemical carcinogenesis-DNA adducts (Figure 7C). Integrated transcriptomic-metabolomic analysis highlighted associations between DEGs and altered metabolites, with joint enrichment in glycolysis/gluconeogenesis, glycerolipid metabolism, cytochrome P450-mediated xenobiotic metabolism, carbon metabolism, purine metabolism, and related carbohydrate metabolic pathways (Figure 7D-E). Correlation analysis identified associations between differentially expressed genes and altered metabolites in the analyzed A549 samples (Figure 7F). These associations do not establish direct regulatory or metabolic relationships.

3. Discussion

AN is a high- volume industrial chemical widely used to manufacture acrylic fibers, ABS/SAN resins, nitrile rubber, and related fine chemical products. Metabolomic studies have investigated potential mechanisms linking volatile organic compound exposure to lung cancer (13). IARC classifies acrylonitrile as carcinogenic to humans (Group 1), with sufficient evidence that it causes lung cancer (14). Experimental studies and mechanistic reviews implicate oxidative stress and oxidative DNA damage in AN toxicity (3, 4).

The molecular events linking AN exposure to lung carcinogenesis remain incompletely defined. From an environmental health perspective, the five-gene signature provides a candidate panel for future investigation of molecular responses potentially associated with AN exposure. Studies combining exposure-characterized human cohorts, established exposure biomarkers such as urinary 2CYEMA, and independent molecular validation are needed to determine whether these genes have utility for exposure-related risk assessment. Nevertheless, the analyses in this study represent different levels of evidence. Network toxicology, machine learning, SHAP analysis, molecular docking/MD simulation, and ssGSEA were used as hypothesis-generating and target-prioritization approaches, whereas A549 transcriptomic-metabolomic profiling provided cell-level experimental evidence of AN-responsive molecular and metabolic alterations. Therefore, the computational findings should be interpreted as candidate mechanistic clues rather than definitive validation, and further exposure-response studies, in vivo models, additional lung cell systems, and population-level cohorts are needed before practical application.

Network toxicology identified candidate pathways shared by predicted AN-associated targets and lung cancer-related genes, suggesting potential molecular links between AN exposure and lung carcinogenesis. These associations are hypothesis-generating and do not establish that AN directly alters these pathways or promotes tumor growth. Differential expression analysis revealed marked transcriptional reprogramming, with upregulated genes mainly associated with cell-cycle progression and autophagy imbalance, including CDC20, UBE2C, and CENPF (16), while increased MIF and

GAPDH suggested inflammatory activation and metabolic rewiring (17, 18). In contrast, downregulated genes were enriched in alveolar epithelial differentiation, surfactant secretion, and lipid metabolism, including SFTPC, CLDN18, and CLEC3B (19-21). These findings raise the possibility that AN-associated lung carcinogenesis results from the convergence of cell-cycle dysregulation, inflammatory chemotactic amplification, immune microenvironment remodeling, and loss of alveolar epithelial identity. Protein-protein interaction analysis identified EGFR as the dominant hub gene, underscoring its central role in oncogenic signaling, tumor cell survival, and immune-contextual regulation (22). EGFR and related signaling mediators regulate receptor tyrosine kinase activity, AP-1-dependent transcription, and chemokine networks, which contribute to lung cancer progression and tumor–microenvironment interactions (23–26). Their identification in the present analysis highlights these pathways as candidate mechanisms potentially linking AN exposure to lung carcinogenesis; however, AN-specific involvement requires experimental validation. GO and KEGG analyses identified enrichment of cytokine–cytokine receptor interaction, TNF, NF-κB, IL-17, and chemokine pathways. Previous studies have examined NF-κB expression in NSCLC (27), shown that IL-17A can promote neutrophil recruitment in an experimental lung tumor model (28), and evaluated IL-8 as a prognostic marker in lung cancer (29) and serum CXCL12 in advanced NSCLC treated with PD-1 inhibitors (30). Because the enrichment analysis used predicted AN targets that overlapped with lung cancer–related genes, whether AN directly activates these pathways or alters immune function remains to be tested. Together with the experimental findings in AN-exposed A549 cells and previous studies of lung cancer, these computational results suggest a candidate framework involving EGFR-related signaling, TNF/NF-κB-associated inflammatory pathways, and cell adhesion and migration. The potential contribution of these pathways to AN-associated lung carcinogenesis requires further investigation, particularly to distinguish AN-specific molecular responses from alterations commonly observed in lung cancer.

Machine learning was further applied to prioritize candidate genes associated with AN-induced lung cancer. LASSO, SVM-RFE, and random forest provided complementary approaches to candidate gene prioritization, while SHAP analysis was used to interpret the selected classifier (15, 31). This strategy prioritized EZH2, PPARG, PARP1, CXCR2, and EGFR as candidate genes at the intersection of predicted AN-associated targets and lung cancer biology. These genes have established roles in epigenetic regulation, lipid metabolism, DNA damage responses, chemokine signaling, and receptor tyrosine kinase activity. Their potential contributions to AN-specific molecular responses remain to be determined. Previous studies have demonstrated the involvement of EZH2 in epithelial–mesenchymal transition and lung cancer progression (32), the context-dependent roles of PPARG in metabolic and inflammatory regulation (33), and the relevance of DNA damage responses, neutrophil-associated immune processes, and EGFR signaling to lung cancer biology (34–36). However, these studies do not establish that the five prioritized genes mediate AN-specific molecular responses.

The use of multiple machine-learning algorithms provided complementary approaches to exploratory candidate gene prioritization. Because the original feature-selection and model-evaluation workflow could not be fully reconstructed, the RF classifier comparison was not considered an independently validated assessment of predictive performance. SHAP analysis was therefore used to characterize the relative contributions of the five genes to the fitted RF model rather than to establish their predictive generalizability or causal involvement in AN-associated carcinogenesis. Independent datasets and a fully documented validation procedure incorporating feature selection within training folds are required to establish the robustness of the proposed signature. SHAP analysis further showed that EZH2 and PARP1 contributed most strongly to model output, followed by PPARG, CXCR2, and EGFR. In the fitted classifiers, higher expression of EZH2, PARP1, and EGFR was associated with tumor

classification, whereas PPARG and CXCR2 showed associations in the opposite direction. These model-derived relationships describe contributions to tumor-versus-normal classification and should not be interpreted as evidence of causal, risk-promoting, or protective biological effects.

The five-gene model yielded descriptive estimates of tumor-versus-normal discrimination in the training and evaluation datasets. Independent validation of the modeling pipeline and assessment of AN-associated lung cancer susceptibility remain separate tasks. Explainable AI methods can characterize how individual features contribute to a model's predictions (37). Molecular docking and dynamics simulations generated structural hypotheses about potential interactions between AN and EZH2 and EGFR. However, AN is a small, reactive chemical, and conventional noncovalent docking does not adequately capture all possible mechanisms of chemical reactivity or protein modification. The predicted binding poses and simulation trajectories therefore require biochemical confirmation and should not be interpreted as evidence that AN directly binds or regulates either protein. Future experiments should distinguish direct chemical interactions from secondary changes in protein expression or activity. AN-derived hemoglobin adducts have been measured as exposure biomarkers (38). Whether AN forms adducts with EZH2 or EGFR requires direct testing. Genetic perturbation could separately test whether either protein contributes to the cellular response to AN. EZH2 has been implicated in drug resistance in NSCLC (39), providing a separate rationale for studying its function in lung cancer cells.

Gene set variation analysis (GSVA) estimated relative gene-set enrichment across samples in the lung cancer transcriptomic dataset (40). These genes were mainly enriched in oncogenic pathways related to inflammation, proliferation, and signal transduction. Positive enrichment of IL-8-CXCR2, JAK–STAT, PI3K-AKT, MAPK, NF-κB, TNF, mTOR, and cell-cycle pathways was consistent with transcriptional programs associated with inflammation and cell proliferation. Conversely, negative enrichment of pathways related to metabolism, ion transport, vesicular trafficking, autophagy, and membrane transport suggested alterations in metabolic and cellular homeostasis. These findings identify candidate molecular processes associated with the analyzed lung cancer transcriptomes but do not establish their activation or suppression by AN exposure (41, 42). Transcriptome-based immune signature analysis further linked these genes to immune-related molecular features in lung cancer. In the analyzed lung cancer dataset, tumor samples exhibited lower ssGSEA scores for several neutrophil-, monocyte-, macrophage-, and dendritic cell-related transcriptional signatures than normal samples, together with a higher chemokine response score. In the analyzed lung cancer dataset, CXCR2 and PPARG generally showed positive correlations with immune-related transcriptional signatures, whereas EGFR, PARP1, and EZH2 showed predominantly negative correlations. These associations may reflect differences in cellular composition, tumor-cell transcriptional programs, or both. Because ssGSEA estimates gene-expression signatures rather than directly measuring immune-cell abundance or function, the findings require validation using orthogonal immune-profiling approaches (43-47). GeneMANIA further showed that the five core genes formed a tightly connected functional module enriched in ERBB signaling, epithelial proliferation regulation, and receptor metabolic processes. In parallel, the miRNA-TF-mRNA network revealed an additional layer of transcriptional and post-transcriptional regulation involving highly interactive modules centered on SP1, BRD4, EP300, and NFKB1, particularly around PPARG and EZH2. Together, these computational analyses identify interconnected signaling, immune-related, metabolic, and regulatory pathways that may be relevant to lung cancer biology. Their specific involvement in molecular responses to AN exposure requires experimental validation.

Integrated transcriptomic and metabolomic analyses suggested that AN exposure induced extensive metabolic remodeling in A549 cells. Metabolic reprogramming is increasingly recognized as a hallmark of environmentally associated lung carcinogenesis. Consistent with recent evidence showing that

particulate matter, polycyclic aromatic hydrocarbons, oxygenated polycyclic aromatic hydrocarbons, and volatile organic compounds can disrupt energy metabolism, oxidative stress responses, and xenobiotic metabolism, our findings indicate that AN also perturbs multiple metabolic pathways. Notably, AN exposure was associated with alterations in glycolysis/gluconeogenesis, glycerolipid metabolism, cytochrome P450-mediated xenobiotic metabolism, glutathione metabolism, fatty acid degradation, and PPAR-related lipid metabolic pathways. This coordinated disturbance of carbohydrate-lipid metabolism, detoxification capacity, oxidative stress-related pathways, and extracellular matrix remodeling may provide candidate markers of AN-responsive metabolic stress in A549 cells. Transcriptomic profiling identified downregulated genes enriched in cytochrome P450-mediated metabolism, glutathione metabolism, fatty acid degradation, and PPAR-related lipid metabolic pathways. Conversely, upregulated genes were enriched in MAPK, PI3K–Akt, PD-L1/PD-1 checkpoint, and extracellular matrix–receptor interaction pathways. These transcriptional changes suggest candidate alterations in xenobiotic metabolism, cellular stress responses, extracellular matrix-related processes, and growth factor-associated signaling. Whether these changes affect detoxification capacity, metabolic activity, or cellular behavior requires direct functional assessment (4, 48, 49). Metabolomic profiling identified changes in metabolites associated with carbohydrate and lipid metabolism, providing complementary evidence of AN-responsive metabolic alterations in A549 cells (33, 50). Joint transcriptomic–metabolomic analysis further highlighted candidate relationships among metabolic pathways, cellular stress responses, and extracellular matrix-associated processes. These findings do not establish altered metabolic flux or demonstrate that the observed changes drive lung carcinogenesis.

Overall, this study integrates network toxicology, machine learning, transcriptomics, metabolomics, and molecular simulations to identify candidate molecular pathways that may link acrylonitrile (AN) exposure to lung cancer. Computational analyses prioritized EZH2, PPARG, PARP1, CXCR2, and EGFR as candidate genes, while experimental profiling identified AN-responsive transcriptional and metabolic changes in A549 cells. Together, these findings generate testable hypotheses concerning inflammatory signaling, cellular stress responses, and metabolic alterations associated with AN exposure. However, they do not establish that the prioritized genes mediate AN-induced lung carcinogenesis or that the proposed five-gene signature is specific to AN exposure. Further validation in nonmalignant pulmonary models, exposure-characterized human cohorts, and mechanistic experiments is required to determine the relevance of these candidate pathways to AN-associated lung cancer.

A central translational challenge is distinguishing molecular features associated with lung cancer from alterations specifically attributable to AN exposure. The computational analyses identify biologically plausible candidate pathways, but many of these pathways are also dysregulated by tobacco smoke, other environmental toxicants, and established oncogenic drivers. Consequently, the five-gene signature may capture general lung cancer biology rather than an AN-specific molecular response. Exposure-stratified human cohorts, appropriate adjustment for smoking and occupational co-exposures, and comparisons with other pulmonary toxicants will be essential to establish specificity. Until such evidence becomes available, the proposed signature should be considered a research tool for target prioritization rather than a validated biomarker for occupational screening or clinical risk prediction.

This study has several limitations. First, the computational analyses integrate predicted chemical targets with lung cancer-associated gene-expression data; the resulting associations do not establish that AN directly regulates the prioritized genes. Second, the original feature-selection and model-evaluation workflow could not be fully reconstructed; consequently, the random forest classifier comparison and SHAP analysis were considered exploratory, and the reported random forest AUC of 1.000 was not

treated as evidence of independently validated predictive performance. The five-gene signature requires evaluation using independent datasets and a fully documented validation procedure that prevents information leakage. Third, docking and molecular dynamics simulations provide structural hypotheses rather than experimental evidence of direct target engagement. Fourth, experimental profiling was restricted to A549 cells exposed to 100 μM AN for 24 h, limiting extrapolation to normal pulmonary epithelium, chronic low-dose exposure, and human occupational conditions. Finally, pathway enrichment and transcriptome-derived immune signatures do not directly measure signaling activity, metabolic flux, or immune-cell function. In addition, the generalizability of the classifier depends on the independence of feature selection and model evaluation, while interpretation of the experimental multi-omics findings requires transparent reporting of biological replication, statistical testing, and multiple-testing procedures. Future studies should incorporate dose–response and time-course experiments, nonmalignant pulmonary models, targeted genetic perturbation, biochemical target validation, and exposure-characterized human cohorts.

4. Materials and Methods

Network Toxicology Analysis of Candidate Genes Linking AN Exposure to Lung Cancer

AN compound identification

We obtained the molecular structure and physicochemical properties of acrylonitrile (AN) from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/; accessed in April 2026). We further evaluated the toxicological profile of AN using multiple platforms, including ADMETLAB 3.0 (https://admetlab3.scbdd.com/; last updated January 31, 2024; accessed in April 2026), ChEMBL 36 (https://www.ebi.ac.uk/chembl/; released October 15, 2025; accessed in April 2026), and ProTox-3.0 (https://tox.charite.de/protox3/; accessed in April 2026).

Identification of AN-associated target genes

Potential human targets of AN were collected by integrating predictions from SwissTargetPrediction (https://www.swisstargetprediction.ch/; 2019 updated model; accessed in April 2026), STITCH v5.0 (http://stitch.embl.de/; last updated January 4, 2016; accessed in April 2026), and TargetNet (http://targetnet.scbdd.com/; accessed in April 2026). The compound name “acrylonitrile” and/or canonical SMILES were used as the search input, and only human targets were retained when species restriction was available. We standardized gene lists from different sources to official gene symbols, then merged and deduplicated them to generate a refined set of AN-related target genes (TRGs).

Lung cancer-associated gene set

Lung cancer-associated genes were retrieved from the GeneCards database (https://www.genecards.org/; accessed in April 2026) and OMIM database (https://www.omim.org/; accessed in April 2026) using the keyword“lung cancer”. The species was restricted to Homo sapiens where applicable. We standardized retrieved genes to official gene symbols, merged them, and deduplicated the list to establish the final lung cancer-related gene set (IRGs).

Acquisition of GEO data and identification of lung cancer-associated differentially expressed genes (DEGs)

The GSE18842 human lung cancer transcriptomic dataset was obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18842; accessed in April 2026). After matching expression data with sample information, we collapsed duplicate gene symbols by mean

expression, imputed missing values using row medians, and normalized the matrix by quantile normalization. We performed differential expression analysis between tumor and normal samples using limma (version 3.62.2) in R version 4.4.3 (2025-02-28 ucrt), with Tumor versus Normal as the contrast. Genes with FDR < 0.05 and |logFC| ≥ 1 were defined as DEGs. We generated volcano plots and heatmaps using ggplot2 (version 4.0.3) and pheatmap (version 1.0.13), respectively, and used the resulting DEGs for downstream analyses.

Identification of AN-Lung Cancer Intersecting Genes (InterGenes)

The intersections among TRGs, LCGs, and DEGs were determined using the ggvenn R package (version 0.1.19). Genes shared by all three datasets were defined as AN-lung cancer-associated intersecting genes (interGenes) and used for subsequent analyses.

Functional Enrichment Analysis

We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of interGenes using the clusterProfiler R package (version 4.14.6), with an unadjusted P < 0.05 threshold for exploratory identification of relevant biological functions and pathways. We obtained KEGG pathway annotations from the KEGG database (https://www.kegg.jp/; accessed in April 2026). We visualized enrichment results using the ggplot2 (version 4.0.3), circlize (version 0.4.18), and ComplexHeatmap (version 2.22.0) R packages.

Protein-protein interactions (PPI) Network Construction and Hub Gene Identification

We analyzed PPIs among interGenes using the STRING database v12.5 (https://string-db.org/; accessed in April 2026), restricted to *Homo sapiens,* with a minimum interaction score of 0.4. We removed disconnected nodes and visualized the resulting PPI network using Cytoscape 3.10.3 (https://cytoscape.org/). We evaluated network topology based on degree centrality and prioritized genes with higher degree values as candidate hub genes.

Machine Learning-Based Prioritization of Candidate Genes

To identify candidate genes associated with AN-related lung cancer, we applied LASSO, SVM-RFE, and random forest to PPI-prioritized candidate genes. GSE18842 was used as the training dataset (n = 91, 46 tumor and 45 normal samples), and GSE32863 was used as the evaluation dataset (n = 116, 58 tumor and 58 normal samples). Stratified five-fold outer resampling was performed with random seed 123. Within each outer split, feature selection and hyperparameter tuning were restricted to the training portion, without using the held-out samples. LASSO used glmnet (version 5.0; binomial family, α=1, standardize=TRUE), with ten-fold inner cross-validation to select λ_{min}, the λ value yielding the minimum mean cross-validated binomial deviance. Genes with non-zero coefficients at λ_{min} were retained and pooled across outer folds. Linear SVM-RFE used caret (version 7.0.1) and kernlab (version 0.9.33), with ten-fold resampling for feature selection and ten-fold cross-validation for model tuning based on ROC AUC. Up to 20 genes were retained by selection frequency across outer folds. Random forest used randomForest (version 4.7.1.2), with 500 trees and other parameters at default settings; up to 20 genes were retained by mean MeanDecreaseGini importance across outer folds. We integrated LASSO-selected features, SVM-RFE feature-selection results, and random forest variable importance to identify candidate genes for subsequent model comparison, SHAP interpretation, logistic regression-based risk modeling, and downstream mechanistic analyses.

SHapley Additive exPlanations (SHAP) Analysis

Eight supervised machine-learning algorithms—DTS, glmBoost, KNN, logistic regression, neural network, PLS, random forest (RF), and SVM—were compared using the five prioritized genes. Because the original feature-selection and model-evaluation workflow could not be fully reconstructed, the classifier comparison was treated as exploratory rather than as an independent assessment of predictive performance. RF was selected for SHAP analysis to examine the relative contributions of individual genes to the fitted model's predictions. SHAP values were used to characterize model behavior and feature importance; they were not interpreted as evidence of causal gene function or AN-specific effects.

Construction and Evaluation of Binary LR Model

We constructed a binary logistic regression model using the five AN-associated genes to calculate individual risk scores. We compared the expression levels of EGFR, PARP1, EZH2, CXCR2, and PPARG between normal and tumor samples. A nomogram was generated using the rms R package (version 8.1.1) to visualize gene contributions to risk prediction. Discrimination between tumor and normal samples was summarized using the area under the receiver operating characteristic curve (AUC) in each dataset. The AUCs were 0.925 in the training dataset and 0.846 in the evaluation dataset. Because the independence of feature selection from model evaluation could not be fully verified, these AUCs are reported as descriptive estimates of discrimination.

Molecular Docking and Molecular Dynamics Simulation

The AN ligand structure was obtained from PubChem (https://pubchem.ncbi.nlm.nih.gov/; accessed in April 2026) and optimized using Open Babel (version 2.4.0) with the MMFF94 force field. Human EZH2 (PDB ID: 5WG6; https://www.rcsb.org/structure/5WG6?utm) and EGFR (PDB ID: 1M17; https://www.rcsb.org/structure/1M17?utm) structures were retrieved from the Protein Data Bank (PDB; PDB; https://www.rcsb.org; accessed in April 2026) and prepared in AutoDock Tools. Molecular docking was performed using AutoDock Vina (version 1.1.2) to predict potential AN–protein binding modes and affinities, and the resulting interactions were visualized using Discovery Studio and PyMOL. Molecular dynamics simulations of the predicted AN–EZH2 and AN-EGFR complexes were subsequently conducted using GROMACS version 2025.4. Following CHARMM36/GenFF2 parameterization, TIP3P solvation, energy minimization, and 2 ns NVT and 2 ns NPT equilibration at 300 K, each complex was subjected to a 100 ns production simulation. We used RMSD, RMSF, Rg, SASA, hydrogen-bond analysis, PCA, and Gibbs free-energy landscape analyses to characterize complex stability and conformational dynamics. These computational analyses were interpreted as evidence of predicted interaction and conformational behavior rather than direct biochemical validation of protein binding or functional regulation.

Immune signature-based analysis

ssGSEA was performed with predefined immune cell and immune-function gene sets to calculate transcriptome-based immune-related signature scores and compare them between normal and tumor groups. Group differences were assessed using two-sided Wilcoxon rank-sum tests and visualized using boxplots. Spearman correlation analysis was used to assess correlations among immune-related signatures and between core genes and ssGSEA-derived scores. An unadjusted $P < 0.05$ threshold was used to identify potentially relevant group differences and correlations for further investigation.

GeneMANIA Functional Prediction and miRNA-TF-mRNA Network Construction

GeneMANIA was used to predict functional modules and interaction networks associated with the core genes. Upstream microRNAs (miRNAs) were identified through integrated interrogation of

miRTarBase, miRWalk, and TargetScan, whereas transcription factors (TFs) were obtained from TRRUST v2. Based on these data, an integrated miRNA-TF-mRNA regulatory network was constructed and visualized using the tidygraph and ggraph packages in R, facilitating the identification of potential regulatory axes and providing mechanistic insights.

Cell Culture

Human lung adenocarcinoma A549 cells were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) and cultured in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin. Cells were maintained at 37 °C in a humidified incubator with 5% CO_2, with medium replaced every 2 days. A549 cells at passages 4-8 after recovery from frozen stocks were used for all experiments. For AN exposure, cells were seeded in 6-well plates at a density of 1×10^5 cells per well and allowed to attach for 24 h. Cells were then treated with Acrylonitrile (AN; CAS No. 107-13-1; Cat. No. A95590; analytical reagent grade; ACMEC Biochemical, Shanghai, China) at a final concentration of 100 μM for 24 h. After treatment, cell lysates were collected for subsequent multi-omics analyses.

Transcriptomic analysis

Three biological replicates per group were analyzed without additional technical replicates. Following RNA quality control, libraries underwent Illumina paired-end sequencing. Filtered reads were aligned using HISAT2 and quantified using featureCounts. Differentially expressed genes were identified using DESeq2 with Benjamini-Hochberg-adjusted $P<0.05$ and $|\log2FoldChange|\geq1.0$.

Metabolomic analysis

Three biological replicates per group were analyzed by UHPLC-MS/MS without additional technical replicates; pooled QC samples were not counted as biological replicates. Data were processed using Compound Discoverer 3.3 with blank-based background removal. Peak areas were divided by the ratio of each sample's total metabolite signal to that of QC1. Metabolites with QC CV>30% were excluded.

Statistical analysis

Differential metabolites were selected using metaX-derived PLS-DA VIP>1, Student's t-test $P<0.05$, and fold change ≥2 or ≤0.5, without additional multiple-testing correction. GO and KEGG enrichment results were screened using unadjusted $P<0.05$ and interpreted as exploratory. Gene-metabolite associations were assessed using Pearson correlation on standardized data from corresponding samples, with $P<0.05$ used to flag associations.

References

1. Wang S, Xing G, Li F, Yang B, Zhang Y, Aschner M, et al. Fasting enhances the acute toxicity of acrylonitrile in mice via induction of CYP2E1. Toxics. 2022;10(6):337. https://doi.org/10.3390/toxics10060337

2. Thier R, Lewalter J, Bolt HM. Species differences in acrylonitrile metabolism and toxicity between experimental animals and humans based on observations in human accidental poisonings. Arch Toxicol. 2000;74(4–5):184–189. https://doi.org/10.1007/s002040000109

3. Pu X, Kamendulis LM, Klaunig JE. Acrylonitrile-induced oxidative stress and oxidative DNA damage in male Sprague-Dawley rats. Toxicol Sci. 2009;111(1):64–71. https://doi.org/10.1093/toxsci/kfp133

4. Albertini RJ, Kirman CR, Strother DE. Acrylonitrile's genotoxicity profile: mutagenicity in search of an underlying molecular mechanism. Crit Rev Toxicol. 2023;53(2):69–116. https://doi.org/10.1080/10408444.2023.2179912

5. Scarselli A, Porzio A, Marinaccio A. Evaluation of occupational exposure risk experienced by a cohort of workers exposed to acrylonitrile using a register-based information system. Ann Work Expo Health. 2025;69(6):602–614. https://doi.org/10.1093/annweh/wxaf021

6. Marsh GM, Youk AO, Collins JJ. Reevaluation of lung cancer risk in the acrylonitrile cohort study of the National Cancer Institute and the National Institute for Occupational Safety and Health. Scand J Work Environ Health. 2001;27(1):5–13. https://doi.org/10.5271/sjweh.581

7. Alexander DD, Pastula ST, Riordan AS. Epidemiology of lung cancer among acrylonitrile-exposed study populations: a meta-analysis. Regul Toxicol Pharmacol. 2021;122:104896. https://doi.org/10.1016/j.yrtph.2021.104896

8. De Jesús VR, Zhang L, Bhandari D, Zhu W, Chang JT, Blount BC. Characterization of acrylonitrile exposure in the United States based on urinary n-acetyl-S-(2-cyanoethyl)-L-cysteine (2CYEMA): NHANES 2011-2016. J Expo Sci Environ Epidemiol. 2021;31(2):377–385. https://doi.org/10.1038/s41370-020-00286-1

9. Keil A, Haber G, Graubard B, Stewart PA, Silverman D, Koutros S. Estimating impacts of reducing acrylonitrile exposure on lung cancer mortality in an occupational cohort with the parametric g-formula. Occup Environ Med. 2024;81(5):225–231. https://doi.org/10.1136/oemed-2024-109532

10. Zhao Q, Zhao Z, Du K. Exploring the association between air pollutants and non-small cell lung cancer using network toxicology and machine learning. Discov Oncol. 2025;16(1):2327. https://doi.org/10.1007/s12672-025-04143-1

11. Wang Z, Wang C, Wan C, Xiang S, Yao J, Sun W, et al. Exploring the mechanisms of Benzo[a]pyrene-induced lung adenocarcinoma based on network toxicology and machine learning. BMC Pharmacol Toxicol. 2025;27(1):19. https://doi.org/10.1186/s40360-025-01064-1

12. Kobets T, Iatropoulos MJ, Williams GM. Acrylonitrile induction of rodent neoplasia: potential mechanism of action and relevance to humans. Toxicol Res Appl. 2022;6:1–33. https://doi.org/10.1177/23978473211055363

13. Yan M, Yang J, Zhu H, Zou Q, Zhao H, Sun H. Volatile organic compound exposure in relation to lung cancer: insights into mechanisms of action through metabolomics. J Hazard Mater. 2024;480:135856. https://doi.org/10.1016/j.jhazmat.2024.135856

14. Stayner LT, Carreón-Valencia T, Demers PA, Fritz JM, Sim MR, Stewart P, et al. Carcinogenicity of talc and acrylonitrile. Lancet Oncol. 2024;25(8):962–963. https://doi.org/10.1016/S1470-2045(24)00384-X

15. Toussaint PA, Leiser F, Thiebes S, Schlesner M, Brors B, Sunyaev A. Explainable artificial intelligence for omics data: a systematic mapping study. Brief Bioinform. 2024;25(1):bbad453. https://doi.org/10.1093/bib/bbad453

16. Zhang S, You X, Zheng Y, Shen Y, Xiong X, Sun Y. The UBE2C/CDH1/DEPTOR axis is an oncogene and tumor suppressor cascade in lung cancer cells. J Clin Invest. 2023;133(4):e162434. https://doi.org/10.1172/JCI162434

17. Youness RA, Elemam NM, Abdelhamid AM, Mohamed AH, Elsherbiny LM, Ramzy A, et al. Macrophage migration inhibitory factor (MIF) and the tumor ecosystem: a tale of inflammation, immune escape, and tumor growth. Front Immunol. 2025;16:1636839. https://doi.org/10.3389/fimmu.2025.1636839

18. Wang K, Li J, Zhang H, Ma H, Cui HY, Ju HQ, et al. Tobacco smoking rewires cell metabolism by inducing GAPDH succinylation to promote lung cancer progression. Cancer Res. 2025;85(15):2838–2857. https://doi.org/10.1158/0008-5472.CAN-24-3525

19. Zuo B, Wang L, Li X, Li X, Wang J, Xiong Y, et al. Abnormal low expression of SFTPC promotes the proliferation of lung adenocarcinoma by enhancing PI3K/AKT/mTOR signaling transduction. Aging (Albany NY). 2023;15(21):12451–12475. https://doi.org/10.18632/aging.205191

20. Sugiyama K, Chau I. Claudins as diagnostic tools and therapeutic targets-glimpse of the horizon. Cancer Treat Rev. 2025;133:102888. https://doi.org/10.1016/j.ctrv.2025.102888

21. Han G, Sinjab A, Rahal Z, Lynch AM, Treekitkarnmongkol W, Liu Y, et al. An atlas of epithelial cell states and plasticity in lung adenocarcinoma. Nature. 2024;627(8004):656–663. https://doi.org/10.1038/s41586-024-07113-9

22. Halder S, Basu S, Lall SP, Ganti AK, Batra SK, Seshacharyulu P. Targeting the EGFR signaling pathway in cancer therapy: what's new in 2023? Expert Opin Ther Targets. 2023;27(4–5):305–324. https://doi.org/10.1080/14728222.2023.2218613

23. Song L, Gong Y, Wang E, Huang J, Li Y. Unraveling the tumor immune microenvironment of lung adenocarcinoma using single-cell RNA sequencing. Ther Adv Med Oncol. 2024;16:17588359231210274. https://doi.org/10.1177/17588359231210274

24. Tan X, Kroneberg M, Sun F, Diaz KA, Flora A, Carey MF. AP-1 promotes oncogenic transcription in lung cancer cells by bridging promoter-enhancer interactions. Cancer Gene Ther. 2026;33(3):301–313. https://doi.org/10.1038/s41417-025-00974-w

25. Jung H, Paust S. Chemokines in the tumor microenvironment: implications for lung cancer and immunotherapy. Front Immunol. 2024;15:1443366. https://doi.org/10.3389/fimmu.2024.1443366

26. Que Z, Xi Z, Qi D, Dai R, Li Y, Liu M, et al. Src/FN1 pathway activation drives tumor cell cluster formation and metastasis in lung cancer: a promising therapeutic target. Sci Adv. 2025;11(28):eadv7377. https://doi.org/10.1126/sciadv.adv7377

27. Zhang L, Ludden CM, Cullen AJ, Tew KD, Branco de Barros AL, Townsend DM. Nuclear factor kappa B expression in non-small cell lung cancer. Biomed Pharmacother. 2023;167:115459. https://doi.org/10.1016/j.biopha.2023.115459

28. Akbay EA, Koyama S, Liu Y, Dries R, Bufe LE, Silkes M, et al. Interleukin-17A promotes lung tumor progression through neutrophil attraction to tumor sites and mediating resistance to PD-1 blockade. J Thorac Oncol. 2017;12(8):1268–1279. https://doi.org/10.1016/j.jtho.2017.04.017

29. Belluomini L, Cesta Incani U, Smimmo A, Avancini A, Sposito M, Insolda J, et al. Prognostic impact of interleukin-8 levels in lung cancer: a meta-analysis and a bioinformatic validation. Lung Cancer. 2024;194:107893. https://doi.org/10.1016/j.lungcan.2024.107893

30. Xu Y, Ding L, Li H, Peng Z, Ding K, Huang Z, et al. Serum cytokine analysis in a cohort of advanced non-small cell lung cancer treated with PD-1 inhibitors reveals predictive markers of CXCL12. Front Immunol. 2023;14:1194123. https://doi.org/10.3389/fimmu.2023.1194123

31. Ahmed Z, Wan S, Zhang F, Zhong W. Artificial intelligence for omics data analysis. BMC Methods. 2024;1:4. https://doi.org/10.1186/s44330-024-00004-5

32. Yang Y, Fan H, Liu H, Lou X, Xiao N, Zhang C, et al. NOP2 facilitates EZH2-mediated epithelial-mesenchymal transition by enhancing EZH2 mRNA stability via m5C methylation in lung cancer progression. Cell Death Dis. 2024;15(7):506. https://doi.org/10.1038/s41419-024-06899-w

33. Zhang J, Tang M, Shang J. PPARγ modulators in lung cancer: molecular mechanisms, clinical prospects, and challenges. Biomolecules. 2024;14(2):190. https://doi.org/10.3390/biom14020190

34. Peng J, Zhou J, Liu G, Kang P, Tang X, Luo Z, et al. Alterations and correlation between DNA damage and repair response and PD-L1 expression in non-small cell lung cancers. BMC Cancer. 2025;25(1):1183. https://doi.org/10.1186/s12885-025-14065-4

35. Hu S, Yan C, Tian Y, Sun W. Neutrophils in non-small cell lung cancer and immunotherapy with PD-1/PD-L1 inhibitors. J Transl Med. 2025;23(1):1313. https://doi.org/10.1186/s12967-025-07084-z

36. Peng L, Li H, Yan W, Liu Y, Zhu H, Hong Y. Integrative profiling of lung cancer biomarkers EGFR, ALK, KRAS, and PD-1 with emphasis on nanomaterials-assisted immunomodulation and targeted therapy. Front Immunol. 2025;16:1649445. https://doi.org/10.3389/fimmu.2025.1649445

37. Karim MR, Islam T, Shajalal M, Beyan O, Lange C, Cochez M, et al. Explainable AI for bioinformatics: methods, tools and applications. Brief Bioinform. 2023;24(5):bbad236. https://doi.org/10.1093/bib/bbad236

38. Osterman-Golkar SM, MacNeela JP, Turner MJ, Walker VE, Swenberg JA, Sumner SJ, et al. Monitoring exposure to acrylonitrile using adducts with N-terminal valine in hemoglobin. Carcinogenesis. 1994;15(12):2701–2707. https://doi.org/10.1093/carcin/15.12.2701

39. Zhang Q, Shi Y, Liu S, Yang W, Chen H, Guo N, et al. EZH2/G9a interact to mediate drug resistance in non-small-cell lung cancer by regulating the SMAD4/ERK/c-Myc signaling axis. Cell Rep. 2024;43(2):113714. https://doi.org/10.1016/j.celrep.2024.113714

40. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. https://doi.org/10.1186/1471-2105-14-7

41. Yi M, Li T, Niu M, Zhang H, Wu Y, Wu K, et al. Targeting cytokine and chemokine signaling pathways for cancer therapy. Signal Transduct Target Ther. 2024;9(1):176. https://doi.org/10.1038/s41392-024-01868-3

42. Qiang M, Chen Z, Liu H, Dong J, Gong K, Zhang X, et al. Targeting the PI3K/AKT/mTOR pathway in lung cancer: mechanisms and therapeutic targeting. Front Pharmacol. 2025;16:1516583. https://doi.org/10.3389/fphar.2025.1516583

43. Han L, Chen Y, Huang N, Zhou X, Lv Y, Li H, et al. Cancer-educated neutrophils promote lung cancer progression via PARP-1-ALOX5-mediated MMP-9 expression. Cancer Biol Med. 2024;21(2):175–192. https://doi.org/10.20892/j.issn.2095-3941.2023.0248

44. Fan K, Zhang BH, Han D, Sun YC. EZH2 as a prognostic-related biomarker in lung adenocarcinoma correlating with cell cycle and immune infiltrates. BMC Bioinformatics. 2023;24(1):149. https://doi.org/10.1186/s12859-023-05271-7

45. Kuhlmann-Hogan A, Cordes T, Xu Z, Kuna RS, Traina KA, Robles-Oteíza C, et al. EGFR-driven lung adenocarcinomas coopt alveolar macrophage metabolism and function to support EGFR signaling and growth. Cancer Discov. 2024;14(3):524–545. https://doi.org/10.1158/2159-8290.CD-23-0434

46. Lin G, Lin L, Chen X, Chen L, Yang J, Chen Y, et al. PPAR-γ/NF-kB/AQP3 axis in M2 macrophage orchestrates lung adenocarcinoma progression by upregulating IL-6. Cell Death Dis. 2024;15(7):532. https://doi.org/10.1038/s41419-024-06919-9

47. Cheng Y, Mo F, Li Q, Han X, Shi H, Chen S, et al. Targeting CXCR2 inhibits the progression of lung cancer and promotes therapeutic effect of cisplatin. Mol Cancer. 2021;20(1):62. https://doi.org/10.1186/s12943-021-01355-1

48. Bae T, Hallis SP, Kwak MK. Hypoxia, oxidative stress, and the interplay of HIFs and NRF2 signaling in cancer. Exp Mol Med. 2024;56(3):501–514. https://doi.org/10.1038/s12276-024-01180-8

49. Bigos KJA, Quiles CG, Lunj S, Smith DJ, Krause M, Troost EGC, et al. Tumour response to hypoxia: understanding the hypoxic tumour microenvironment to improve treatment outcome in solid tumours. Front Oncol. 2024;14:1331355. https://doi.org/10.3389/fonc.2024.1331355

50. Li B, Wang D, Wang Y, Huang Z, Liu Q, Zheng Z, et al. Distinct metabolic profiles in lung adenocarcinomas presenting as solid or ground-glass opacities. NPJ Precis Oncol. 2026;10(1):174. https://doi.org/10.1038/s41698-026-01378-1

Declarations

Funding Statement: This work was supported by the Research Program of Tianjin Municipal Education Commission (grant number 2025ZXZD009).

Competing Interests: The authors declare no financial or personal relationships with other individuals or organizations that could inappropriately influence or bias the content of this work. All authors have read the final version of the manuscript and confirm that there are no competing interests.

Consent for publication: All authors have approved the final version of the manuscript.

Use of Artificial Intelligence Disclosure: This article was written by human contributors. Artificial intelligence-based tools were used to improve grammar, language, and readability without affecting the article's scientific content, data interpretation, or conclusions. The authors reviewed and verified all content to ensure its accuracy and integrity.

Data Availability Statement: The publicly available lung cancer transcriptomic datasets GSE18842 and GSE32863 were obtained from the NCBI Gene Expression Omnibus (GEO). The transcriptomic and metabolomic datasets generated from the A549 cell experiments in this study have been deposited

in the China National GeneBank Sequence Archive (CNSA) of CNGBdb under project accession numbers CNP0010302 and CNP0010308, respectively. Additional data supporting the findings of this study are provided in the article and its Supplementary Material.

Supplementary Material:
https://cancerbiometherapy.com/index.php/cbtt/libraryFiles/downloadPublic/42

Ethics approval and consent to participate: Not applicable. This study analyzed publicly available, de-identified transcriptomic data and conducted experiments using an established human lung cancer cell line. No new human participants or animal experiments were involved.

Authors' affiliations
1. State Key Laboratory of Advanced Medical Materials and Devices, School of Medicine, Tianjin University, Tianjin, China.

2. Department of Clinical Laboratory, Lequn Branch, The First Hospital of Jilin University, Changchun, Jilin Province

3. Department of Gastroenterology and Hepatology, Changi General Hospital, Singapore, Singapore.

4. SingHealth Duke-NUS Academic Medical Centre, Singapore 169857, Singapore.

5. Key Laboratory of Medical Rescue Key Technology and Equipment, Ministry of Emergency Management, School of Disaster and Emergency Medicine, Tianjin University, Tianjin, China.

6. Wenzhou Safety (Emergency) Institute of Tianjin University, Wenzhou, China.

7. Department of Respiratory Medicine, Haihe Hospital, Tianjin University, Tianjin, China.

CRediT authorship contribution statement: WL, YH, CZ, TC, and HY contributed to writing the original draft, methodology, data curation, and figure preparation. LW and YL contributed to manuscript review, critical revision, and editing. JD and YC contributed to conceptualization and supervision. JD contributed to conceptualization, supervision, and funding acquisition. All authors contributed to preparing their respective sections and revising the complete manuscript. All authors read and approved the final manuscript.

ORCID ID
Wenjing Li: https://orcid.org/0009-0003-4224-6893
Yaqi Hu: https://orcid.org/0009-0006-2558-9933
Chaofan Zhang: https://orcid.org/0009-0008-4021-7606
Tiansong Cao: https://orcid.org/0009-0009-3326-4604
Louis Zizhao Wang: https://orcid.org/0000-0001-8350-5220
Yanqing Liu: https://orcid.org/0000-0002-2270-096X
Hongzhi Yu: https://orcid.org/0000-0001-5004-2370
Yang Cao: https://orcid.org/0009-0003-9301-040X
Jinrui Dong: https://orcid.org/0009-0000-8377-4973