



OPEN ACCESS

EDITED BY

Zhenwei Shi,
Guangdong Academy of Medical Sciences,
China

REVIEWED BY

Han Zhang,
Tianjin University, China
Yiluo Xie,
Bengbu Medical College, China

*CORRESPONDENCE

Hongmei Zhang
✉ zhanghm@ncst.edu.cn
Yi Guan
✉ gy110901@163.com

RECEIVED 02 September 2025

REVISED 14 December 2025

ACCEPTED 25 December 2025

PUBLISHED 16 January 2026

CITATION

Li A, Cui P, Liu L, Liu J, Zhou X, Wu W, Yan Z,
Guan Y and Zhang H (2026) Multi-omics
integration and machine learning identify
NPC2 as a prognostic and treatment-
responsive regulator in lung adenocarcinoma.
Front. Immunol. 16:1697560.
doi: 10.3389/fimmu.2025.1697560

COPYRIGHT

© 2026 Li, Cui, Liu, Liu, Zhou, Wu, Yan, Guan
and Zhang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Multi-omics integration and machine learning identify NPC2 as a prognostic and treatment-responsive regulator in lung adenocarcinoma

Ang Li^{1,2}, Ping Cui², Lili Liu³, Jiawei Liu⁴, Xianlei Zhou¹,
Wenlong Wu⁵, Zimo Yan⁵, Yi Guan^{1*} and Hongmei Zhang^{1,6*}

¹School of Public Health, North China University of Science and Technology, Tangshan, China,

²Department of Public Health, Jining Medical University, Jining, Shandong, China, ³Healthcare Associated Infection Control Department, Affiliated Hospital of Jining Medical University, Jining, China, ⁴Department of Immunology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, ⁵College of Life Science, North China University of Science and Technology, Tangshan, China, ⁶School of Clinical Medicine, North China University of Science and Technology, Tangshan, China

Background: This study aims to define a novel molecular subtype of LUAD by integrating multiple omics data. Additionally, we develop and validate an Artificial Intelligence Derived Prognostic Index (AIDPI) that predicts the prognosis of LUAD patients, identifies potential therapeutic targets.

Methods: This study employed ten clustering algorithms from the R package “MOVICS” to integrate multi-omics data of LUAD sourced from TCGA database for molecular typing. Subsequently, an Artificial Intelligence Derived Prognostic Index (AIDPI) was constructed as the most effective indicator for predicting the overall survival rate of LUAD patients. The biological functions and mechanisms of NPC2 in lung adenocarcinoma were elucidated through both *in vitro* and *in vivo* experiments, which included CCK-8 assays, colony formation assays, flow cytometry, Transwell assays, and xenograft tumor models. Additionally, the impact of NPC2 on Ribociclib sensitivity was investigated through drug correlation analysis and molecular docking, while the predictive value of NPC2 regarding immunotherapy benefits was validated using the immune cell infiltration analysis.

Results: Through multi-omics clustering, we identified two subtypes of lung adenocarcinoma associated with prognosis, with the CS1 subtype exhibiting the most favorable prognostic outcomes. The low AIDPI group exhibited a more positive prognosis, accompanied by increased immune cell infiltration and activation of immune pathways. Meanwhile, NPC2 was recognized as a standalone risk factor for LUAD, with its high expression significantly improving the overall survival of LUAD patients. Functionally, the overexpression of NPC2 promotes tumorigenesis in LUAD both *in vitro* and *in vivo*. Mechanistically, the upregulation of NPC2 expression inhibits the progression of LUAD by suppressing the PI3K/AKT signaling pathway. Our study also demonstrated that high NPC2 expression is positively correlated with Ribociclib sensitivity, as confirmed by *in vitro* experiments. Furthermore, NPC2 expression is positively correlated with ImmuneScore, and may serve as a predictive indicator for the efficacy of immune checkpoint inhibitor (ICI) therapy.

Conclusion: The comprehensive analysis of multiple omics data significantly enhances the molecular classification of lung adenocarcinoma. Furthermore, AIDPI is a potential biomarker that predicts the prognosis of LUAD patients. NPC2 inhibits the progression of LUAD by suppressing the PI3K/AKT signaling pathway and enhancing the chemotherapy sensitivity to Ribociclib.

KEYWORDS

innate immunity, lung adenocarcinoma, machine learning, multi-omics, NPC2

1 Introduction

Adenocarcinoma of the lung represents the most prevalent pathological type among non-small cell lung cancers (1), with over 60% of patients diagnosed at stage III or even stage IV. Despite advancements in diagnostic techniques and treatment modalities, the inherent high heterogeneity and drug resistance associated with lung cancer contribute to a dismal 5-year survival rate of less than 15% (2, 3). Recently, candidate biomarkers for the clinical diagnosis of patients have emerged, including PD-L1 expression, TMB, neoantigen burden, mismatch repair defects, and high microsatellite instability. However, the utility of these methods is constrained by spatiotemporal heterogeneity, moderate accuracy, and their applicability to only a small percentage of the patient population (4–6). Consequently, in the context of personalized treatment, it is imperative to identify reliable biomarkers to enhance the prognosis of LUAD and optimize the benefits of pharmacotherapy.

Recent studies indicate that alterations in epigenetics, typically induced by prolonged inflammation, can regulate various facets of cancer advancement, such as tumor occurrence, development, metastasis, and immune evasion (7–9). Toll-like receptors (TLRs) represent a significant category of pattern recognition receptors that play a critical role in bridging innate and adaptive immunity while regulating inflammatory responses (10). Recent research has shown that TLRs mediate the inflammatory response within the tumor microenvironment (TME) in a complex manner. Pathogen-associated molecular patterns (PAMPs) are identified and bound by TLRs, which then recruit and activate downstream signaling molecules, including MyD88 or TRIF. This activation leads to the engagement of transcription factors like IRF3 or NF- κ B, induce the production of type I interferons and inflammatory mediators, and participate in immune inflammatory responses (11, 12). As the most extensively studied pattern recognition receptors, TLR activation promotes dendritic cells maturation and cytokine production, recruits CD8⁺ T cells, and activates specific T cell immune responses, thereby inhibiting tumor occurrence and progression (13, 14). Previous reports have indicated that the regulation of TLR activation is closely associated with the reactivation of tumor-infiltrating lymphocytes' cytotoxic activity (15, 16), and when

combined with chemotherapy or immunotherapy, it can also inhibit tumor growth (17). Therefore, an in-depth exploration of the impact of the TLR pathway on tumor cells and its potential clinical significance is of great scientific importance.

The high intertumor and intratumor heterogeneity of lung adenocarcinoma can result in significantly different clinical outcomes for patients at the same stage (18, 19). In clinical settings, the processes of making decisions, managing treatments, and conducting follow-ups continue to depend on conventional TNM staging systems that are based on anatomical features, which do not take molecular biological features into account. A suitable biomarker ought to demonstrate consistent expression both internally and externally within tumor tissues to operate efficiently for all patients. Consequently, multi-gene expression analysis may represent an encouraging strategy to tackle this variability (20). Recent advancements in multi-omics integration and machine learning have revolutionized the molecular stratification of cancers, offering new avenues for prognostic modeling and therapeutic target identification. For instance, Chu et al. employed ten multi-omics clustering algorithms to identify consensus molecular subtypes in muscle-invasive urothelial cancer, demonstrating that integrative approaches enhance prognostic accuracy and reveal subtype-specific therapeutic vulnerabilities (21). Similarly, in lung adenocarcinoma and clear cell renal cell carcinoma, similar frameworks were used to construct molecular subtypes or prognostic models, effectively predicting the prognosis and immune therapy response of patients (22, 23). These studies underscore the potential of multi-omics consensus clustering and machine learning in uncovering biologically relevant subtypes and prognostic biomarkers.

This study developed integrated common subtypes of LUAD, combining mRNA transcriptome, genomic mutation, and epigenetic DNA methylation data of Toll-like receptor pathway genes. Subsequently, various machine learning algorithms were utilized to construct the Artificial Intelligence Derived Prognostic Index (AIDPI) and identify the key gene NPC2. The NPC2 protein, a small glycoprotein that is soluble, plays a crucial role in binding free cholesterol and regulates its transport and homeostasis within cells (24). Further exploration was conducted on the biological roles and molecular processes of NPC2 in LUAD, offering a foundational

framework and new directions for the prevention and treatment of this disease.

2 Material and methods

2.1 Acquisition and preprocessing of multi omics data

This study collected comprehensive mRNA transcriptome expression profiles, DNA methylation status, somatic mutation profiles, and clinical characteristic data of lung adenocarcinoma patients from the Cancer Genome Atlas (TCGA) database. The gene symbols and names of mRNA were annotated using the GENCODE 27 annotation file. The beta values of DNA methylation were obtained from the Illumina DNA Methylation 450 platform via the UCSC Xena website. Somatic mutations were acquired through cBioPortal (<https://www.cbioportal.org/>) and processed using the maftools software package. After merging all available omics and clinical features, a total of 429 lung adenocarcinoma patients were included in this study for further analysis.

2.2 Acquisition and preprocessing of external validation data

To validate the new findings from multiple omics cohorts, this study selected three external Gene Expression Omnibus (GEO) cohorts, namely GSE72094, GSE13213, and GSE50081, from the GEO database. The mRNA expression matrices and clinical data were obtained from these cohorts. To ensure accuracy and consistency, probes corresponding to multiple molecules were initially removed, retaining only the probe data with the highest signal intensity for each molecule. All chip expression profile data underwent background correction using the limma software package, followed by log2 transformation and normalization. **Table 1** presents a detailed list of the basic feature information of the dataset.

2.3 Multi omics comprehensive clustering analysis

The core functions of the MOVICS software package consist of three modules: the GET module is responsible for the

comprehensive clustering of multiple omics data to identify molecular subtypes; the COMP module is utilized for multidimensional comparisons of features across different subtypes; and the RUN module performs tag recognition and subtype verification tasks. This study employed the “MOVICS” multi-omics integrated analysis R software package to integrate various omics data, including the mRNA transcriptome expression profile of the Toll like receptor signaling pathway gene set, all methylation probe data obtained the UCSC platform (Illumina DNA Methylation 450 chip), and somatic mutation profiles (missense, nonsense, splice site mutations, etc.) from whole exome or whole genome sequencing data, to identify molecular subtypes of lung adenocarcinoma. Initially, the TPM expression data was assessed using univariate Cox regression analysis after log2 transformation, with a screening threshold set at $P < 0.05$. To ascertain the optimal number of subtypes, the “getClustNum” function was employed, in conjunction with multiple omics data, to calculate the clustering prediction index and gap statistic. Subsequently, based on the identified subtypes, patients were classified using ten clustering algorithms, including iClusterBayes, moCluster, CIMLR, IntNMF, ConsensusClustering, COCA, NEMO, PINSplus, and SNF, with LRA as an additional method. By employing a consistency ensemble algorithm, the clustering results were integrated to ensure that the same sample was categorized consistently across different algorithms. Finally, to account for potential discrepancies in omics data, the “getConsensus MOIC” function was utilized to calculate the consistency matrix, quantify the robustness of sample clustering, screen for highly robust subtypes, and evaluate the similarity of samples within subtypes through contour scoring.

2.4 Establish a comprehensive machine learning driven prognostic model

Based on transcriptome data from The Cancer Genome Atlas Lung Adenocarcinoma (TCGA- LUAD), the “limma” package was utilized to analyze differentially expressed genes (DEGs) between normal and tumor tissues, as well as among different molecular subtypes. The screening criteria were established at $|\log_2FC| > 2$ and $FDR < 0.05$. Additionally, the “VennDiagram” package was employed to visualize the intersecting DEGs between the two groups. Subsequently, univariate Cox regression analysis was conducted to identify genes with potential prognostic significance for further investigation. This study utilized TCGA-LUAD data as the model training set and selected three independent lung

TABLE 1 Basic information of GEO LUAD datasets.

GEO datasets	Platform	Disease	Sample		Years	Country
			Control	Case		
GSE72094	GPL15048	LUAD	0	442	2018	United States
GSE13213	GPL6480	LUAD	0	117	2019	Japan
GSE50081	GPL570	LUAD	0	127	2019	Canada

adenocarcinoma cohorts (GSE72094, GSE13213, and GSE50081) from the GEO database as external validation sets. To construct the optimal prognostic model, we integrated ten machine learning methods, including CoxBoost, stepwise Cox regression, Lasso regression, ridge regression, elastic net, support vector machine, generalized additive model, supervised principal component analysis, partial least squares regression, and random survival forest, resulting in a total of 101 algorithm combinations. Each algorithm combination model was validated on both internal and external validation sets, and its robustness and clinical applicability were assessed using the concordance index (C-index). According to the median AIDPI for each group, patients with lung adenocarcinoma were divided into high AIDPI and low AIDPI categories. The prognostic value was analyzed through Kaplan-Meier curves, and Receiver Operating Characteristic (ROC) curves were created to evaluate the prognostic efficacy of AIDPI.

2.5 Gene set variation analysis and gene set enrichment analysis

The estimation of a collection of 50 hallmark gene sets (h.all.v7.5.symbols.gmt) obtained from the MsigDB database was performed using GSVA and GSEA. The “ComplexHeatmap” R package was utilized to visualize significant pathways across various groups.

2.6 TME analysis

To assess the abundance of cells within the tumor microenvironment (TME), we analyzed the expression of NPC2 in individual cells using the TISCH2 database (<http://tisch.comp-genomics.org/home/>). Then, we evaluated the presence of 24 different immune cell types in the TME using Gene Set Variation Analysis (GSVA). An analysis of the distribution of immune checkpoints was carried out, along with the determination of immune and stromal scores for tumor tissues through the “ESTIMATE” package in R. These correlation results were visualized using heatmaps.

2.7 Survival analysis

This study obtained three clinical prognostic endpoint data from the TCGA database: overall survival (OS), progression free interval (PFI), and disease-specific survival (DSS), to comprehensively evaluate the predictive performance of the model. (1) OS: refers to the time from the diagnosis of a patient to death for any reason. (2) PFI: refers to the time from diagnosis to the first occurrence of any of the following events (primary tumor progression, local recurrence, distant metastasis, or death from any cause). (3) DSS: refers to the time from diagnosis to direct death due

to the disease. Due to varying definitions of endpoints, there are inherent differences in sample sizes and event counts across the analyzed cohorts. Survival analysis was visualized using the Kaplan-Meier method, and between-group comparisons were performed with the log-rank test. All analyses were carried out using the “Survival” package and the “Survminer” package in R software.

2.8 Quantify the immunotherapy response predictor

The Tumor Immune Dysfunction and Exclusion (TIDE) framework serves as a tool to assess the likelihood of tumor immune evasion by analyzing tumor gene expression data. We analyzed mRNA-seq data from the TCGA-LUAD, GSE50081, GSE72094, and GSE13213 employing the TIDE algorithms to forecast how patients might respond to Immune-Checkpoint Blocker (ICB) therapy.

2.9 Construction of nomogram

Nomogram plots were constructed by combining the NPC2 expression and various clinical data. In the nomogram graph, every factor was associated with a specific score, which contributed to the cumulative score. The overall scores were then utilized to predict the 1-year, 3-year, and 5-year overall survival (OS) rates; a higher score correlated with a decreased survival rate.

2.10 Cell culture

The human LUAD cell lines A549 and PC9 were provided by Pricella Biotechnology (Wuhan, China) and maintained in RPMI 1640 supplemented with 10% FBS and 0.1 mg/mL penicillin-streptomycin under regular conditions (5% CO₂, 37 °C).

2.11 CCK-8 assay

Cell growth was assessed with a CCK-8 kit (Dojindo, Kumamoto, Japan) based on the manufacturer’s recommendations. A549 and PC9 cells were seeded into 96-well plates, and incubated for 24, 48, and 72 hours. Next, cells were incubated for 1h after adding with 10μL CCK-8 solution. Eventually, the OD450nm was measured using a microplate reader (BioTek, Winooski, VT, USA).

2.12 Clone formation

1×10³ cells were plated into a 35 mm dish and maintained at 37° C and 5% CO₂ for 14 days or observed under microscope when the

number of cells in a single cell mass exceeded 50. Afterwards, the cells were fixed with 4% paraformaldehyde and stain with 0.1% crystal violet (Solarbio, Beijing, China) to count colony numbers.

2.13 Cell migration and invasion assay

Cell migration and invasion was evaluated by a transwell assay. Tumor cells were inoculated into the upper chamber coated with or without Matrigel (Corning Inc., Corning, NY, USA). The lower chamber was filled with RPMI 1640 and 10%FBS. After one day's incubation, cells in the lower chamber were stained with 0.2% crystal violet for 20min in the dark, and stained cells were captured under a microscope (Olympus, Tokyo, Japan).

2.14 Real-time fluorescence quantitative PCR

RNA was extracted from tissues or cells with Trizol reagent (Thermo Fisher Scientific, MA, USA) and RNA reverse transcription was performed using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, MA, USA). After this step, use SYBRGreen Mix (Mei5 Biotechnology, Beijing, CHN) for RT-qPCR. The $2^{-\Delta\Delta Ct}$ method was used, GAPDH was applied as an internal control for normalizing. The primer sequences utilized in the experiment are shown in [Supplementary Table 1](#).

2.15 Apoptosis assay

FITC-Annexin V and propidium iodide were used for double staining in accordance with the Annexin V-FITC/PI Cell Apoptosis Detection Kit (Servicebio, Wuhan, China) manufacturer's instructions, followed by flow cytometry. Flow cytometry data were imported and analysed using the flowCore and flowGate packages.

2.16 Western blot analysis

Lysed cells with RIPA buffer (Thermo Fisher Scientific, MA, USA), and the concentration of total protein was measured by BCA kit (Solarbio, Beijing, China). It was electrophoresed with 10% SDS-PAGE gel, then transferred to nitrocellulose film for 70 minutes at 200mA, sealed with 5% skim milk powder for 1 hour, then incubated overnight at 4°C in primary antibody, washed with TBST and incubated in the HRP-conjugated second antibody at room temperature for 1 hour, then washed with TBST for three times, each time for 10 minutes, and Immunoreactive proteins were detected by enhanced chemiluminescence reagent (Pierce, Thermo Scientific) using a Bioimaging System. The antibodies used are as follows: Bax (ab32503), Bcl-2 (ab182858), E-Cadherin (ab231303), N-Cadherin (ab245117), PI3K (ab191606), p-AKT (phospho T308) (ab38449) and p-PI3K (phospho Y464) (ab138364) were purchased from Abcam (Eugene, USA). β -actin was used as a control.

2.17 Immunohistochemical analysis

Paraffin was used to embed tissue samples, and sliced thinly colored with hematoxylin and eosin dyes (C0105M, Beyotime, China). For immunohistochemistry, the sections were incubated with primary antibodies at 4°C overnight. Subsequently, the sample underwent washing with PBS and incubated with the secondary antibody at 37°C for 2h. After being washed and dehydrated with gradient ethanol and treated with xylene in turn before being prepared for microscopic observation. The mouse antibodies used are as follows: Ki-67 (GB14102) antibodies were purchased from Servicebio (Wuhan, China).

2.18 Xenograft models

Male BALB/c nude mice at four weeks of age were randomly assigned to two groups, with six mice in each group. A549 cells (3×10^6 /0.2ml PBS) stably transfected with NPC2 and empty vector were subcutaneously injected into nude mice to establish xenograft model. Xenografts were examined every 3 days with digital calipers. After 23 days, the mice were sacrificed and stripped the tumors for further analysis. All mice were bought from SPF Biotechnology (Beijing, China) and used in accordance with the agreement approved by the Animal Ethics and Welfare Committee of North China University of Science and Technology (2025-SY-3048).

2.19 Drug sensitivity analysis

The Genomics of Drug Sensitivity in Cancer (GDSC) database comprises a vast collection of genomic datasets and drug sensitivity information. We utilized the “pRRopheticPredict” R package (version 0.5) to construct statistical models leveraging gene expression data derived from a large array of cancer cell lines. Subsequently, we applied these models to the gene expression data of the target samples to assess the variations in IC50 values of 198 drugs between the different NPC2 expression.

2.20 Molecular docking

The interaction between Ribociclib and NPC2 was examined through molecular docking techniques. The 3D structure of Ribociclib was obtained from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and subsequently converted into PDB format utilizing Open Babel 2.3.2 software. The 3D structure of the NPC2 protein was retrieved from the protein database and downloaded in PDB format. The ligand and water molecules were eliminated using this software. Next, PYMOL 2.3.4 was employed to modify the NPC2 protein by adding hydrogen atoms, balancing the charge, and performing other necessary operations through Auto DockTools, which produced PDB files. These were then converted into pdbqt format for subsequent processing. The docking

simulations were conducted using AutoDock Vina 1.1.2 software, and the results were visualized with PYMOL.

2.21 Statistical analysis

The data are presented as mean $\pm$ standard deviation (SD). To assess the differential expression between the two groups, the Student's *t* test was utilized. Survival analysis was conducted using the Kaplan-Meier method and assessed with the log-rank test. All statistical evaluations were carried out with R software (version 4.2.3), SPSS 23.0 (Chicago, IL), and GraphPad Prism 6 (La Jolla, CA, USA). A *P*-value of less than 0.05 was deemed statistically significant.

3 Results

3.1 Multi omics integrated consensus subtypes of lung adenocarcinoma

Based on the Toll like receptor pathway, we found that according to the results of CPI analysis and gap statistical analysis, the highest average statistical value of subtype number was defined as the two clustering results, and the scores of the two analysis methods were more similar (Figure 1A). Furthermore, we utilized 10 multi omics ensemble clustering algorithms to classify patients (Figure 1B), and aggregated them through consistent classification to obtain a more stable clustering result (Figure 1C). Finally, two new molecular subtypes of lung adenocarcinoma were named CS1 and CS2, respectively. Patients with the CS2 subtype experienced the poorest clinical outcomes, in contrast to those with the CS1 subtype, who exhibited the longest overall survival times ($P < 0.03$, Figure 1D). Utilizing multi-omics data from the TCGA-LUAD cohort, the varying distributions of mRNA expression levels, DNA methylation at CpG sites, and mutant genes were subsequently illustrated through a consensus ensemble approach (Figure 1E).

3.2 Development and validation of the AIDPI

We developed the AIDPI following the schematic in Supplementary Figure S1A. Initially, TCGA-LUAD cohort was employed as a training set. Meanwhile, GSE72094, GSE13213 and GSE50081 served as our validation set. Based on the expression profiles of 235 CS1 vs CS2 difference gene expression, univariate Cox analysis identified 47 prognostic mRNA and shared between the training and validation sets (Supplementary Figures S1B-C). The CPGs were incorporated into a machine-learning framework to develop 101 different prediction models using the TCGA-LUAD dataset. Subsequently, the C-index of each model was computed across all validation datasets. The model created through the combination of RSF was selected as the best model since it had

the highest average C-index of 0.713 (Figure 2A). Using this model, we computed the AIDPI for each patient within various cohorts. According to the optimal threshold of the AIDPI established from the training set, LUAD patients were divided into low AIDPI and high AIDPI groups. Furthermore, to more rigorously validate our model, we assessed the predictive performance of AIDPS in the validation cohorts. Kaplan-Meier survival analysis indicated that the low AIDPS group experienced significantly longer overall survival (OS) in the three external validation cohorts (log-rank $P < 0.001$ in the GSE72094, $P = 0.039$ in the GSE13213, $P = 0.013$ in the GSE50081) (Figure 2B). The ROC analyses revealed that the areas under the ROC curves (AUCs) in the training set (TCGA) were 0.979, 0.982 and 0.974 for 1-, 3- and 5-year OS, respectively. The AUCs in the GSE72094 validation set were 0.652, 0.633 and 0.834, GSE13213 validation set were 0.801, 0.715 and 0.648, GSE50081 validation set were 0.575, 0.665 and 0.668 (Figure 2C). Overall, the AIDPS demonstrated a reliable and precise ability to forecast the outcomes for patients with PACA, indicating that it could emerge as a valuable resource for clinical application.

3.3 Prediction of clinical features and immune processes based on AIDPI

Next, we will compare the clinical information and pathological feature distribution of LUAD patients in high- and low- AIDPI groups and sought to elucidate relationships between these factors and the AIDPI. Result revealed significant associations between high- and low- AIDPI groups and parameters such as T stage ($\chi^2 = 13.011$, $P = 0.001$), M stage ($\chi^2 = 4.798$, $P = 0.028$), pathologic stage ($\chi^2 = 10.605$, $P = 0.001$) and radiotherapy ($\chi^2 = 15.501$, $P = 0.001$) (Table 2). In summary, these results suggest that AIDPI may play an important role in tumor metastasis, disease progression, and radiotherapy response of lung adenocarcinoma, and may become a potential biomarker for evaluating the condition and treatment strategies of LUAD patients. Further analysis of multiple independent datasets (GSE13213, GSE50081, and GSE72094) revealed that AIDPI levels in patients with advanced-stage (III-IV) lung adenocarcinoma were significantly elevated compared to those in early-stage (I-II) patients ($P < 0.05$) (Figure 3A). Additionally, AIDPI levels were found to be lower in patients with EGFR mutations and higher in patients with TP53 mutations, as indicated by the GSE13213 and GSE72094 datasets ($P < 0.05$) (Figure 3B).

To further investigate the molecular characteristics between the high and low AIDPI groups, we selected the Hallmark gene set and quantified it using the GSVA algorithm. The results indicated that, compared to other pathways, the low AIDPI group was predominantly enriched in immune-related processes such as coagulation, complement activation, interferon- α response, and IL-6/JAK/STAT3 signaling. In contrast, the high AIDPI group exhibited enrichment mainly in proliferation-related processes, including G2M checkpoints, mitotic spindles, and E2F targets (Figure 3D). Given the significant differences in immune processes between the high and low AIDPI groups, we quantified

the infiltration levels of immune cells in the microenvironment along with the expression of immune checkpoints. As illustrated in the Figure 3C, although the high AIDPI group demonstrated greater activation of CD4⁺ memory T cell infiltration, the low AIDPI group

appeared to exhibit higher immune infiltration activity, characterized by the activation of plasma cells and CD8⁺ T cells, as well as substantial NK cell infiltration. Furthermore, in the low AIDPI group, anti-inflammatory immune cells such as Treg cells,

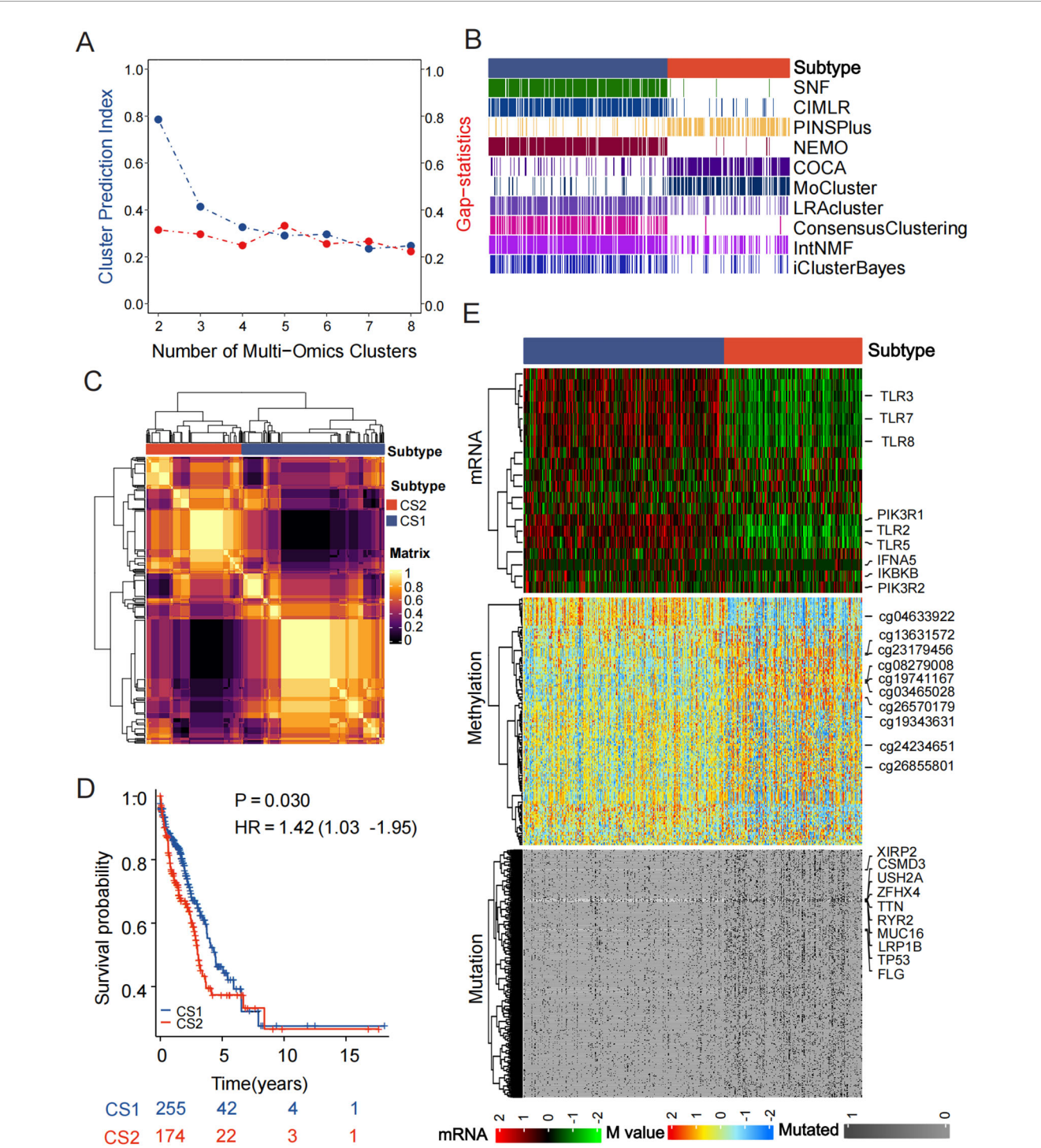


FIGURE 1
The multiomics integrative consensus subtypes of LUAD. **(A)** Cluster prediction indicators and Gap statistics predict the optimal number of clusters for multi omics clustering. **(B)** Clustering of LUAD patients through 10 cutting-edge multiomics clustering methods. **(C)** A consensus clustering matrix showcasing three new prognostic subtypes derived from the ten algorithms. **(D)** Varied survival outcomes observed between the two subtypes. **(E)** Comprehensive heatmap of consensus ensemble subtypes.

follicular helper T cells, and M2 macrophages were significantly more prevalent. Notably, we observed significant differences in immune checkpoint expression between the two groups, with PD-1, PD-L1, CTLA-4, and TIGIT expression levels being markedly

higher in the low AIDPI group than in the high AIDPI group. The distinct patterns of immune infiltration and immune-related gene expression among AIDPI patients in different groups suggest that the response to immunotherapy may vary.

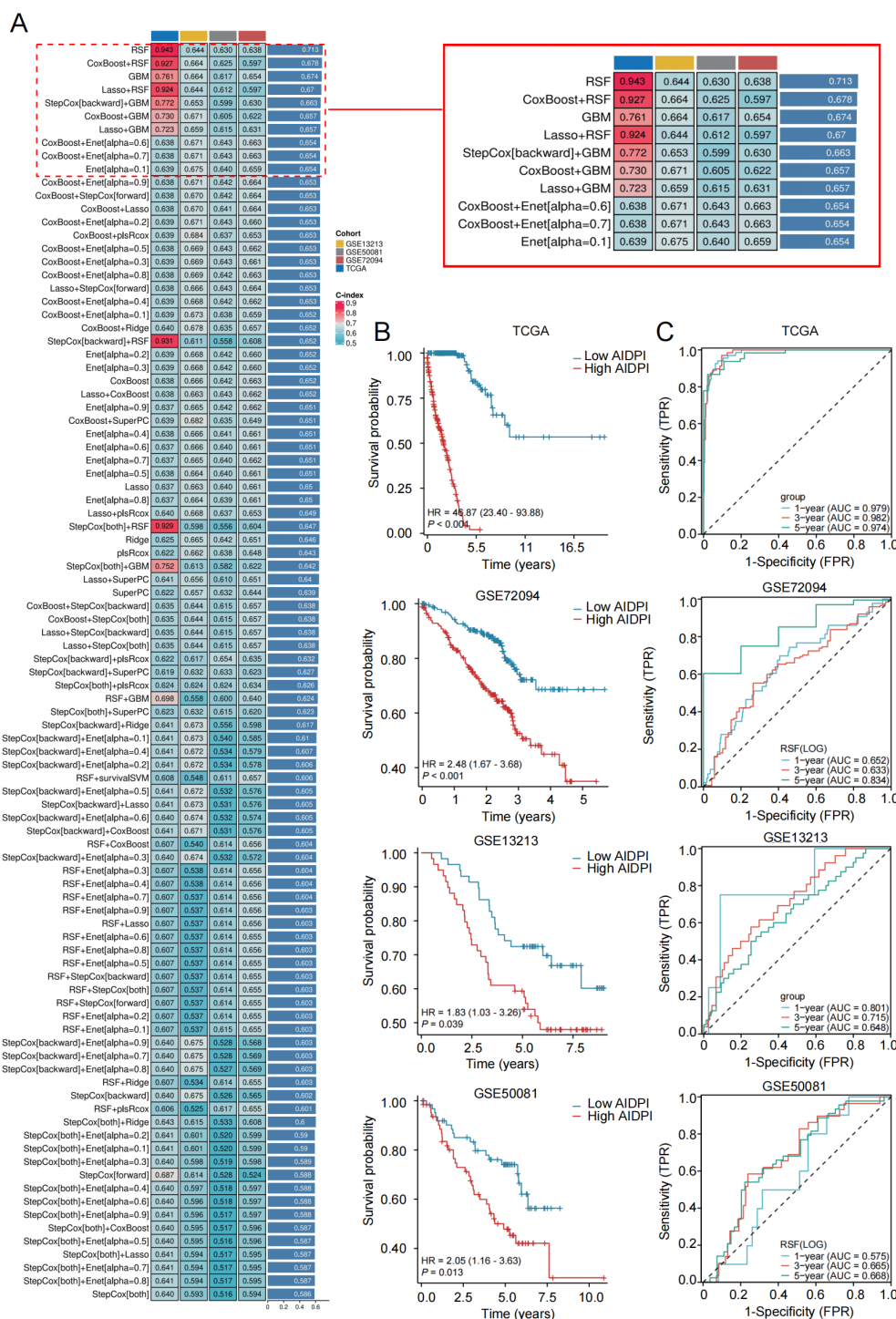


FIGURE 2

Development and validation of the AIDPI. (A) C-indices obtained from different combinations of machine-learning algorithms across three cohorts. (B) Kaplan-Meier survival analyses of the AIDPI across multiple cohorts. (C) Time-dependent ROC curve assessments of the AIDPI performed in several cohorts.

TABLE 2 The relationship between AIDPI level and clinicopathological characteristics of patients with lung adenocarcinoma in TCGA database.

Clinical characteristics	Number	The level of AIDPI		χ^2	P value
		Low (n=215)	High (n=214)		
Age				0.602	0.438
≤60	275	141 (66.5%)	134 (62.9%)		
>60	150	71 (33.5%)	79 (37.1%)		
Gender				1.735	0.188
Male	196	91 (42.9%)	105 (49.3%)		
Female	229	121 (57.1%)	108 (50.7%)		
Smoking				1.644	0.199
Yes	366	178 (84.0%)	188 (88.3%)		
No	59	34 (16.0%)	25 (11.7%)		
T Stage				3.573	0.059
T1&T2	372	192 (90.6%)	180 (84.5%)		
T3&T4	53	20 (9.4%)	33 (15.5%)		
N Stage				13.011	0.001
N0	280	156 (75.4%)	124 (58.8%)		
N1&N2&N3	138	51 (24.6%)	87 (41.2%)		
M Stage				4.798	0.028
M0	270	136 (96.5%)	134 (89.9%)		
M1	20	5 (3.5%)	15 (10.1%)		
Pathological Stage				10.605	0.001
Stage I & Stage II	283	157 (74.1%)	126 (59.2%)		
Stage III & Stage IV	142	55 (25.9%)	87 (40.8%)		
Radiotherapy				15.501	0.001
No	367	197 (92.9%)	170 (79.8%)		
Yes	58	15 (7.1%)	43 (20.2%)		
Chemotherapy				0.785	0.376
No	280	144 (67.9%)	136 (63.8%)		
Yes	145	68 (32.1%)	77 (36.2%)		

3.4 NPC2 low expression in LUAD correlates with tumor progression

We optimized the key parameters of the random forest through cross validation, particularly mtry (reducing the number of randomly selected features per tree), nodesize (increasing the minimum sample size of leaf nodes), and nsplit (limiting the number of node partitions), to enhance the regularization effect. Further assessed genes based on their significance (nrep=1000, which denotes that the Monte Carlo simulation was run for 1000 iterations; nstep=5, meaning that five steps were progressed) (Figure 4A). The connection between the error rate and the count of classification trees is depicted in Figure 4B. In the end, genes

demonstrating a relative importance exceeding 0.6 were chosen as possible candidates. Through additional intersection analysis alongside differentially expressed genes from high and low AIDPI groups, we identified three overlapping genes. The expression patterns of these genes were later depicted using feature maps (Figure 4C). According to the canSAR database (<https://cansar.ai/>), only the protein encoded by CYP4B1 and NPC2 possesses a druggable structure, making it a potential target for patients with high AIDPI. In addition, there have been numerous literature reports that CYP4B1 is a prognostic biomarker and potential therapeutic target for lung adenocarcinoma, and is associated with immune cell infiltration (25, 26). Therefore, we have identified NPC2 as the main gene for further research. Our

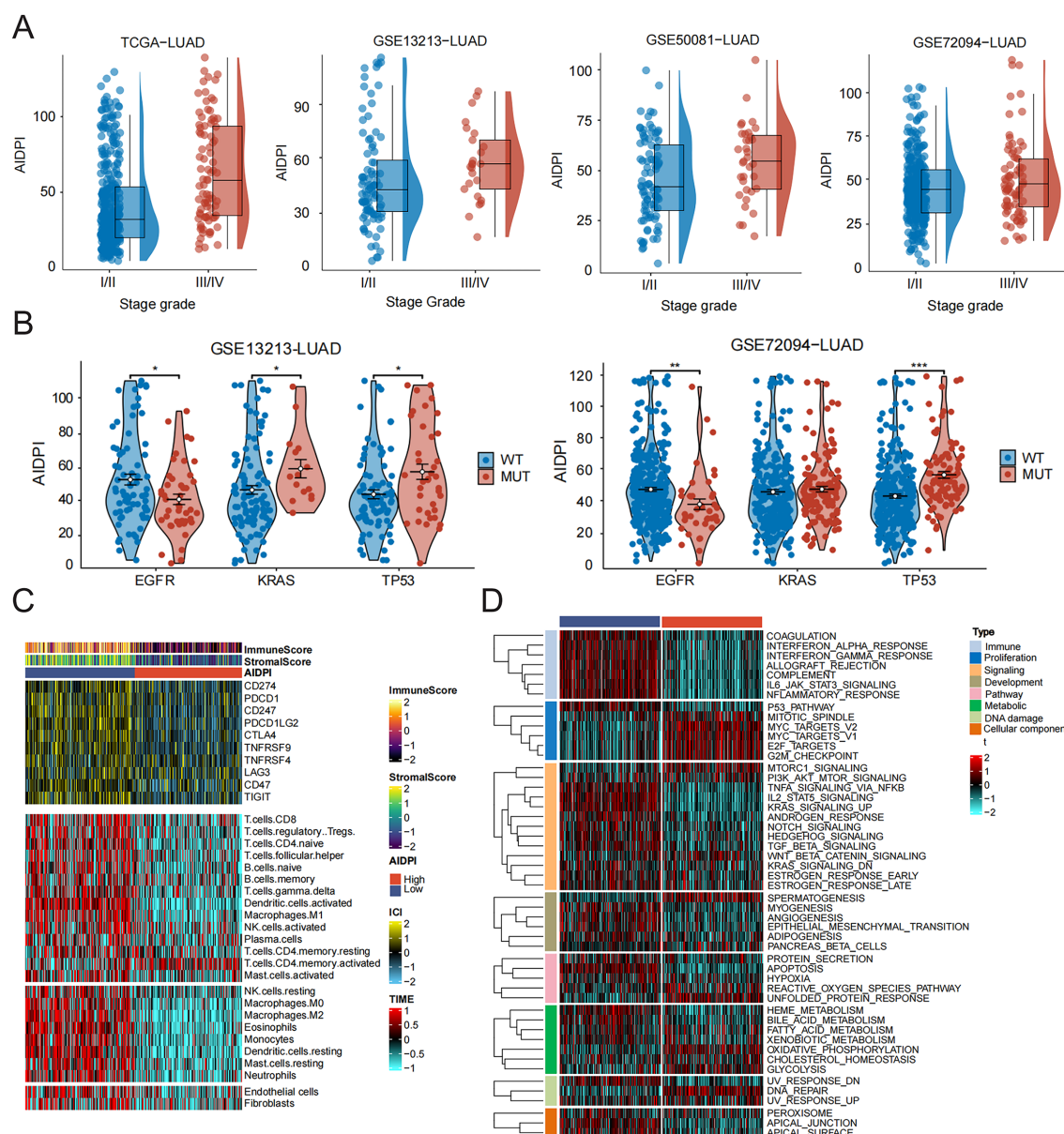


FIGURE 3
Prediction of clinical features and immune processes based on AIDPI. **(A)** AIDPI distributions across varying pathological staging in the indicated cohorts. **(B)** AIDPI distributions across GFR, KRAS, and TP53 mutations patient in the indicated cohorts. **(C)** Expression of immune checkpoints and the enrichment of immune cell infiltration within the TCGA cohort. **(D)** Signal pathway enrichment analysis of 50 hallmark gene sets.

analysis of data from TCGA-LUAD reveals that the expression level of NPC2 is significantly reduced in LUAD tissues compared to normal tissues (Figure 4D). Follow-up research investigating the clinical results of NPC2 and LUAD within three groups (TCGA, GSE50081, and GSE13213) revealed that reduced expression of NPC2 correlates with unfavorable prognosis in patients with LUAD, supported by survival curves from various cohorts (Figures 4E–I). This indicates that patients in the low NPC2 group exhibit significantly lower overall survival (OS) or event-free survival rates across various LUAD cohorts. Further analysis revealed that NPC2 mRNA levels in late-stage (III–IV) lung

adenocarcinoma patients were significantly higher than those in early-stage (I–II) patients ($P < 0.05$) (Figures 4J–L).

Considering the acknowledged influence of various clinical factors on the outcomes for LUAD patients, we aim to elucidate the relationship between these factors and NPC2 to validate NPC2 as an independent prognostic biomarker. The multidimensional circular plot illustrating the survival status of patients with varying clinical characteristics in the NPC2 high and low expression groups demonstrates that the survival rate of patients in the NPC2 low expression group is significantly higher than that of the high expression group across several dimensions, including overall

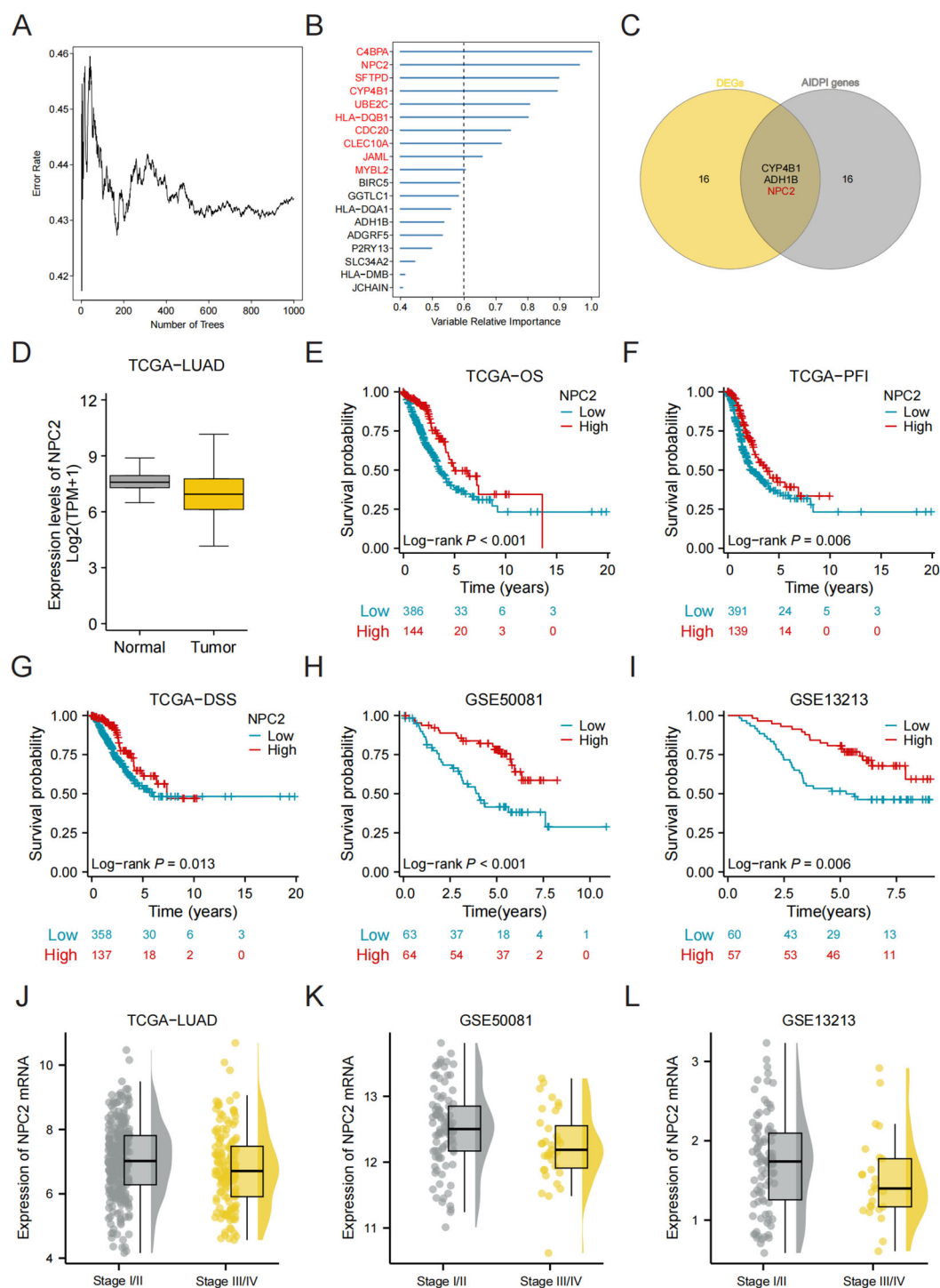


FIGURE 4

High NPC2 expression is associated with good prognosis in LUAD. (A) The error rate related to the data according to the classification tree. (B) Screening of biomarkers utilizing the RF model. (C) A Venn diagram illustrates the common entities across the specified gene groups. (D) Analysis of NPC2 expression in both normal tissue and LUAD through the TCGA dataset. (E–I) Kaplan-Meier plots for NPC2 expression derived from the TCGA, GSE13213 and GSE50081 datasets. (J–L) Distribution of NPC2 expression in early and late stage LUAD patients.

survival (OS), gender, tumor stage, primary tumor focus, lymph node metastasis, distant metastasis, and age (Figure 5A). Univariate and multivariate Cox regression analyses demonstrated a significant correlation between overall survival (OS) and NPC2 expression, T

stage, N stage, and pathological stage in patients with lung adenocarcinoma (Figures 5B, C). When assessing its predictive ability in smokers, NPC2 mRNA expression was found to be lowest in smoking patients and highest in non-smoking lung

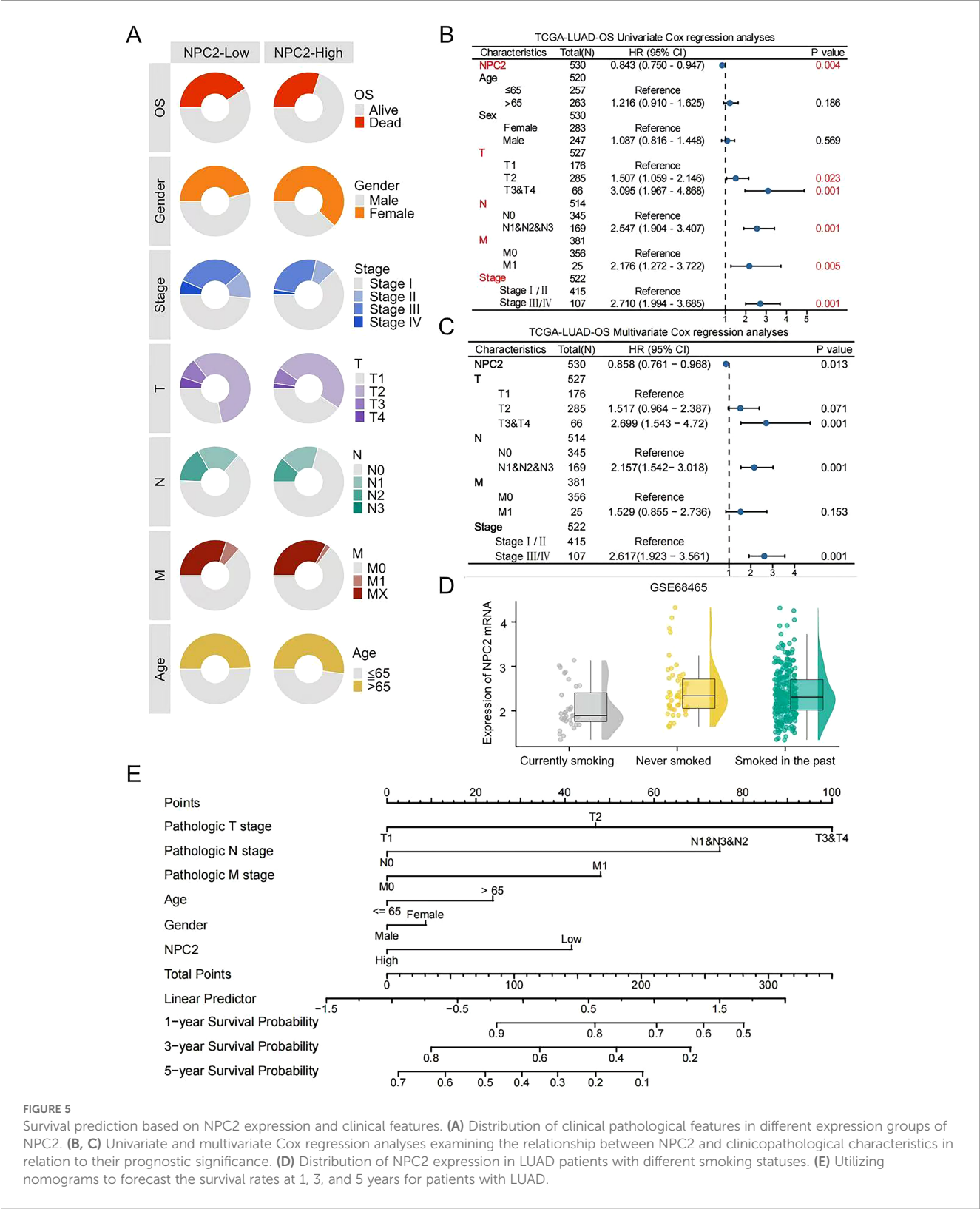


FIGURE 5 Survival prediction based on NPC2 expression and clinical features. **(A)** Distribution of clinical pathological features in different expression groups of NPC2. **(B, C)** Univariate and multivariate Cox regression analyses examining the relationship between NPC2 and clinicopathological characteristics in relation to their prognostic significance. **(D)** Distribution of NPC2 expression in LUAD patients with different smoking statuses. **(E)** Utilizing nomograms to forecast the survival rates at 1, 3, and 5 years for patients with LUAD.

adenocarcinoma patients (Figure 5D). Based on these findings, we constructed a nomogram model to estimate the effect of NPC2 expression on the OS of lung adenocarcinoma patients at 1, 3, and 5 years. This model incorporates age, gender, TNM staging, and NPC2 mRNA expression, as illustrated in Figure 5E. Overall, our results indicate that the column chart model based on NPC2 demonstrates strong predictive performance for the prognosis of LUAD patients.

3.5 NPC2 overexpression impedes LUAD *in vitro* and *in vivo*

The Human Protein Atlas database indicates low NPC2 protein abundance in LUAD tissues by representative immunohistochemical images (Figure 6A). To clarify the role of NPC2 in cell biological behaviors, we established stable A549 and PC9 cell through lentiviral-mediated NPC2 overexpression or control with verified the overexpression efficiency (Figure 6B). Function assays using these cell lines revealed that NPC2 overexpression significantly suppressed colony formation and proliferation of LUAD cells (Figures 6C, D). Moreover to gain insights into the effect of NPC2 on LUAD metastasis, transwell migration/invasion assays were performed (Figure 6E). The results showed that NPC2 suppressed LUAD cell metastasis *in vitro*. The transition from epithelial to mesenchymal cells (EMT) is an essential characteristic associated with the invasion and metastasis of tumors. Consequently, western blot assays were employed to analyze markers related to EMT. The findings indicated that the overexpression of NPC2 enhanced the levels of E-cadherin while reducing the levels of N-cadherin (Figure 6F). To elucidate the mechanism underlying NPC2 induction of LUAD cell proliferation, cell apoptosis were further detected by flow cytometry. The results verified that NPC2 overexpression increased the rate of apoptotic cells (Figure 6G). Consistent with the flow cytometry results, NPC2 overexpression significantly downregulated apoptosis inhibitory protein Bcl-2, and upregulated the expression of proapoptotic proteins Bax (Figure 6H). Taken together, these data provided evidences that NPC2 suppressed LUAD cell proliferation, promoted cell apoptosis, and cell metastasis *in vitro*.

3.6 NPC2 overexpression inhibits PI3K/AKT pathway and suppresses *in vivo* tumor growth in LUAD

In order to uncover intracellular pathways that could facilitate the influence of NPC2 within LUAD cells, we conducted a bioinformatics analysis of related genes derived from the TCGA-LUAD dataset. Hallmark pathway analysis showed that NPC2 was linked to the PI3K/AKT pathway (Figure 7A). Western blot analysis of A549 and PC9 cells demonstrated a significant decrease in the expression levels of the active, phosphorylated (p) forms of AKT and PI3K proteins in LUAD cells that overexpress NPC2 (Figure 7B). Thus, we proposed that NPC2 could influence the progression of LUAD through the inhibition of the PI3K/AKT signaling pathway. Rescue experiments demonstrated that both overexpression NPC2 and the PI3K activator 740 Y-P promoted the NPC2 mediated decreased proliferation, migration, and invasion capacity of LUAD cells (Figures 7C, D). These results indicated that NPC2 inhibits cell proliferation and migration/invasion by suppressing the AKT/mTOR signaling pathway in LUAD cells.

To investigate the roles of NPC2 in LUAD tumorigenesis *in vivo*, A549 cells with upregulated NPC2 were subcutaneously

injected into BALB/c mice. The average volume and weight of tumors in the groups with NPC2 overexpression were significantly reduced compared to the control group (Figures 7E–G). Immunohistochemical (IHC) analysis revealed a notable reduction in Ki-67 expression in the xenograft tumor tissues of the NPC2 overexpression groups when juxtaposed with those from the control tumors (Figure 7H). These results indicated that overexpression NPC2 suppressed the growth of LUAD cells *in vivo*, which was consistent with the results *in vitro*.

3.7 Pharmacologic NPC2 overexpression enhances chemotherapy efficacy

The observed suppressive effects of NPC2 overexpression on LUAD led us to explore its therapeutic potential further. We initially performed a drug-correlation analysis against NPC2 expression. The expression of NPC2 mRNA is negatively correlated with Ribociclib sensitivity, meaning that as NPC2 expression increases, the IC₅₀ of Ribociclib decreases (Figure 8A). Research shows that Ribociclib is a cyclin dependent kinase inhibitor. Subsequently, GSEA pathway analysis was conducted based on the HALLMARK and KEGG pathway gene sets, and the results showed that high expression of NPC2 was associated with inhibition of the cell cycle pathway (Figure 8B). To determine whether elevated levels of NPC2 enhance sensitivity to Ribociclib, further investigation through molecular docking analysis was prompted. The results from the molecular docking indicated a robust and stable interaction between Ribociclib and NPC2, featuring a binding energy of -7.8 kcal/mol. Notable docking sites were identified, with the residue GLN-387 involved in interactions through hydrogen bonding (Figure 8C). These findings indicating NPC2 potential as a therapeutic agent in modulating Ribociclib drug sensitivity in LUAD treatment. Finally, functional validation was undertaken using colony formation assays. These results suggest that NPC2 may potentiate the inhibitory effect of Ribociclib on LUAD cell proliferation (Figure 8D). These findings indicate that upregulation of NPC2 enhances sensitize LUAD cells to Ribociclib, supporting its potential as a therapeutic strategy for lung adenocarcinoma.

Additionally, to further elucidate the impact of NPC2 on immunotherapy, we analyzed data from the TCGA-LUAD, GSE13213, and GSE50081 datasets. The results of the ESTIMATE analysis indicated a positive correlation between NPC2 expression and ImmuneScore ($P < 0.001$, $r = 0.347$; $P < 0.001$, $r = 0.425$; $P < 0.001$, $r = 0.191$) (Supplementary Figure S2A). Utilizing the TIDE algorithm to predict potential immune responses, we found that NPC2 expression was significantly elevated in responding patients compared to non-responding patients (Supplementary Figure S2B). Meanwhile, NPC2 expression exhibited the strongest correlation with the infiltration of T follicular helper cells and mast cells (Supplementary Figure S2C). Single cell transcriptome analysis based on non-small cell lung cancer datasets (GSE131907 and

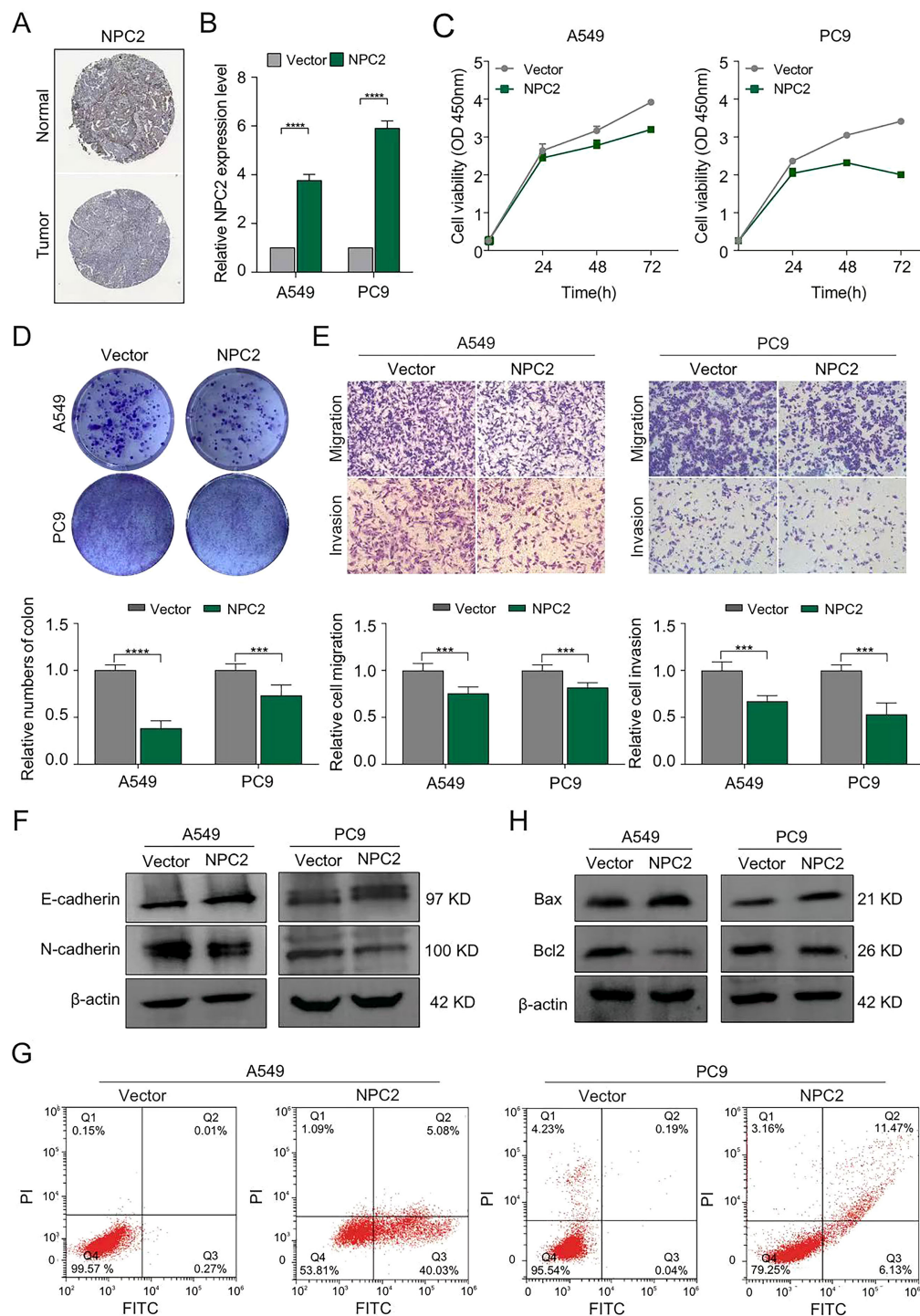


FIGURE 6

NPC2 inhibits tumor growth and metastasis of lung adenocarcinoma. **(A)** Analysis of protein expression differences of NPC2 in lung adenocarcinoma tissue and normal tissue through HPA database. **(B)** qRT-PCR detection of overexpression efficiency of NPC2. **(C)** CCK-8 assay in LUAD cells stably up-regulating NPC2 on cell viability. **(D)** Colony formation assays in LUAD cells stably up-regulating NPC2. **(E)** The transwell assays were utilized to assess the migration and invasion of the specified LUAD cells. **(F)** The epithelial mesenchymal transition related proteins E-cad and N-cad were detected by western blot following NPC2 overexpression. **(G)** Apoptosis related proteins were detected after NPC2 overexpression. **(H)** The analysis of cell apoptosis resulting from the overexpression of NPC2 was conducted using flow cytometric assays. The data are the means $\pm$ SEM. *** $P < 0.001$, **** $P < 0.0001$.

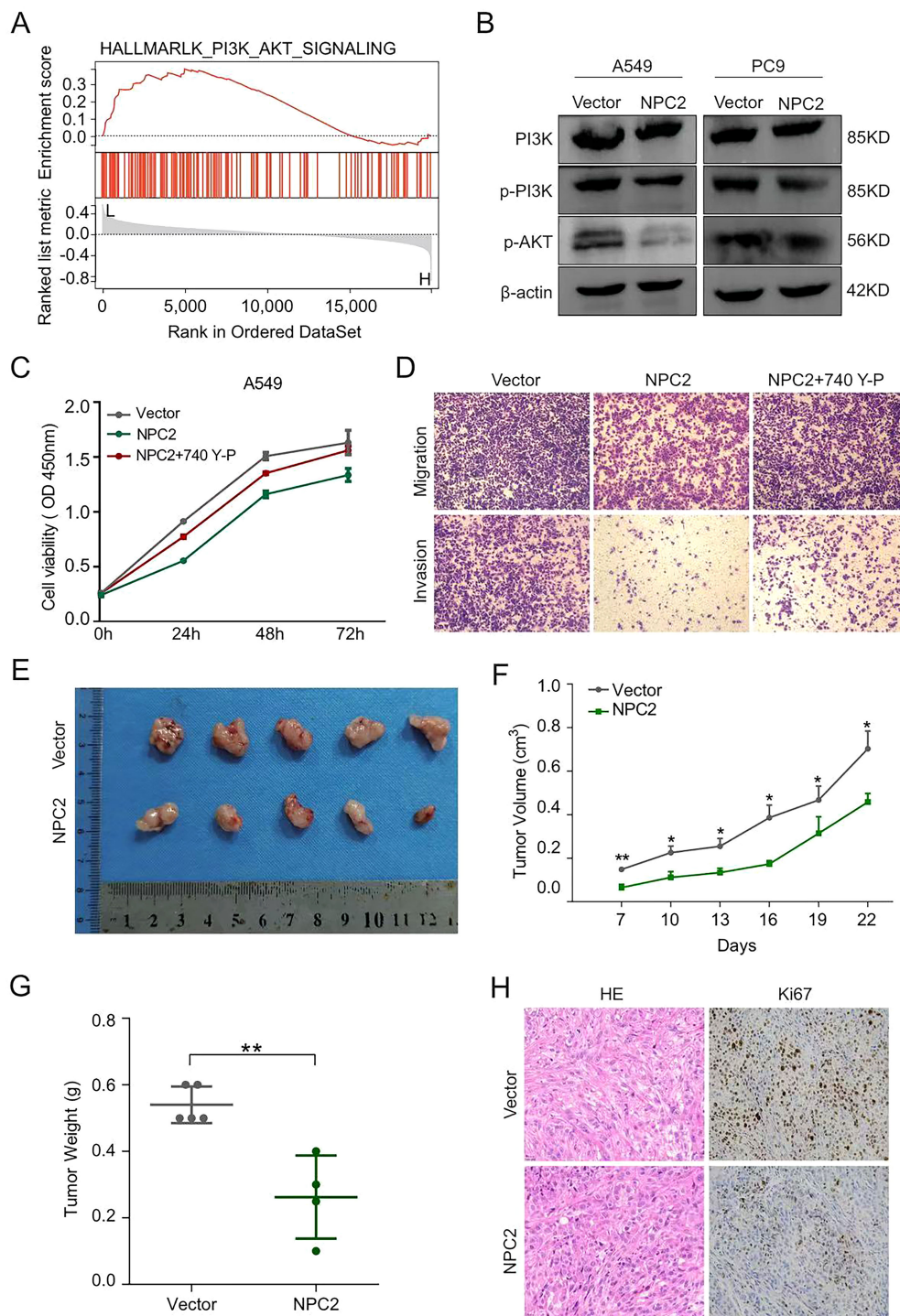


FIGURE 7

NPC2 overexpression inhibits PI3K/AKT pathway and suppresses *in vivo* tumor growth in LUAD. (A) GSEA results were illustrated to demonstrate the association of NPC2 with the PI3K/AKT signaling pathway. (B) Evaluate the expression levels of crucial proteins within the PI3K/AKT pathway in LUAD cells. (C) CCK-8 assays were used to observe the growth of LUAD cells. (D) Cell migration and invasion ability were observed using the transwell assay. (E) Representative image of tumors formed in nude mice. (F) Growth curves showing the subcutaneous tumor volume across various groups. (G) Subcutaneous weight of tumor in nude mice. (H) Representative images for IHC staining of tumor samples from the different groups. ** $P < 0.01$.

EMTAB6149) showed that NPC2 is mainly expressed in monocytes/macrophages, epithelial and fibroblasts (Supplementary Figure S2D). Collectively, these findings provide preliminary support for the predictive significance of NPC2 in immune checkpoint inhibitor

(ICI) immunotherapy. In the future, further *in vivo* and *in vitro* models are necessary to investigate the synergistic effects of NPC2 overexpression in conjunction with immune checkpoint inhibitors for the treatment of LUAD.

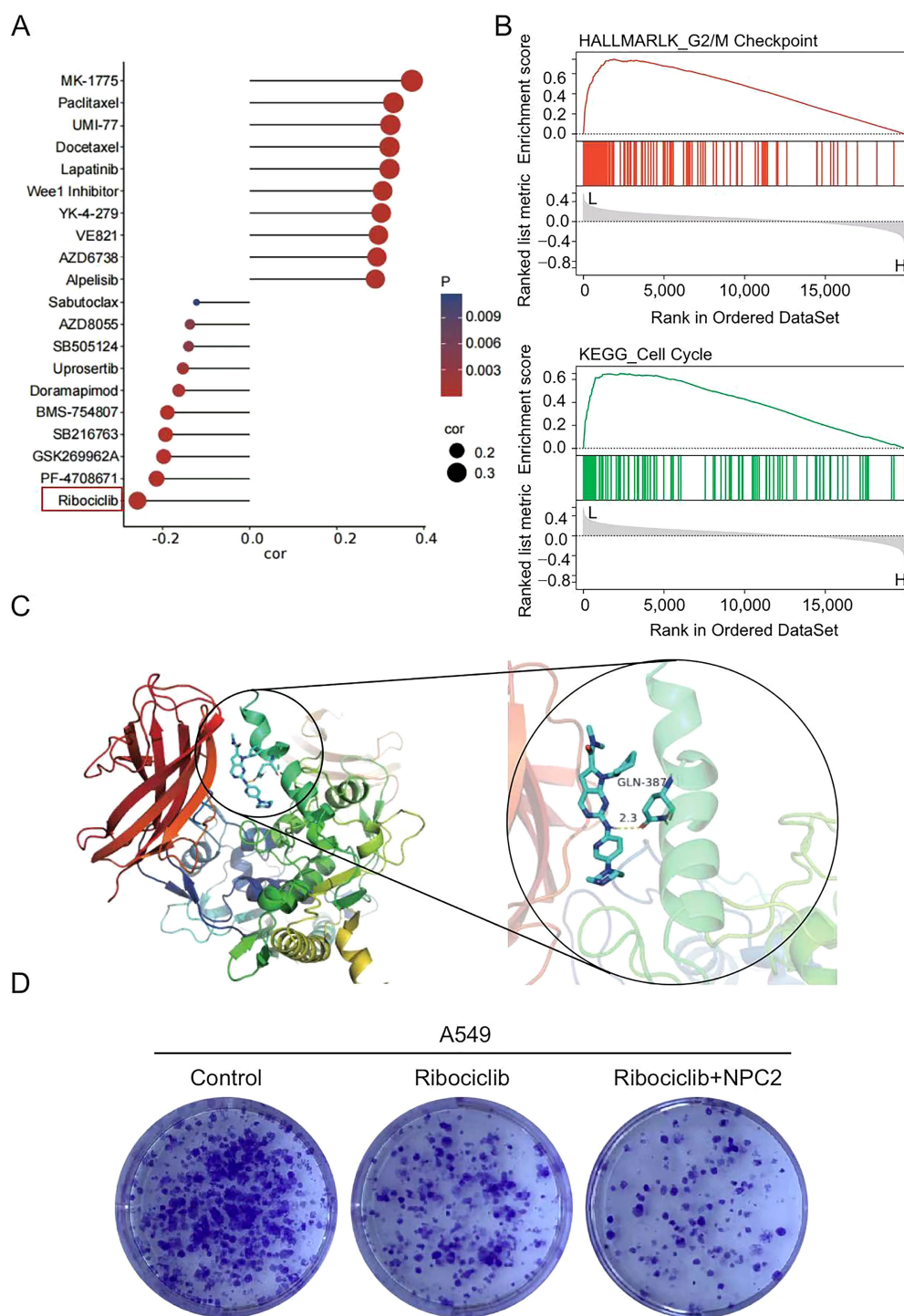


FIGURE 8

Pharmacologic NPC2 overexpression enhances chemotherapy efficacy. (A) Analyzing the correlation between the expression level of NPC2 and sensitivity to drugs. (B) GSEA analysis of the effect of NPC2 expression on cell cycle pathways. (C) Molecular docking analysis. (D) Proliferative ability of A549 cells was measured by colony formation assay.

4 Discussion

Lung cancer remains the predominant cause of cancer-related deaths worldwide, with lung adenocarcinoma (LUAD) being the most common

histological variant, making up roughly 50% of all cases of non-small cell lung cancer (NSCLC) (27, 28). Despite recent advancements in early diagnosis, molecular classification, and immunotherapy, the 5 year survival rate for LUAD patients remains low. Consequently, it is

essential to clarify the molecular processes that drive the progression of LUAD in order to discover new biomarkers and therapeutic targets.

Gene expression is meticulously regulated by several epigenetic mechanisms, including methylation, mutation, and histone modification (29). Consequently, the integrated analysis of multi-omics data holds significant potential for screening tumor markers and identifying pathogenic targets, as well as facilitating early diagnosis and treatment (30). Currently, numerous studies have focused on the joint analysis of multiple omics. For instance, consensus immune-related lncRNA features, constructed through machine learning ensemble methods, can serve as indicators to evaluate prognosis, recurrence, and treatment benefits for colorectal cancer patients receiving fluorouracil-based ACT, bevacizumab, and pembrolizumab (31). Chen et al. employed LASSO analysis and 10-fold cross-validation to identify 22 image factors associated with bladder cancer, highlighting the potential clinical applications of machine learning in analyzing bladder cancer histopathological images (32). The application of random forests in machine learning to develop tumor microenvironment phenotypic signatures for breast cancer, based on the TME phenotypic characteristics of imageomics, can assist in predicting patients' clinical responses to anti-PD-1/PD-L1 treatment (33). However, given that these prognostic models rely on expression profiles of mRNAs, miRNAs, or lncRNAs within particular pathways such as immunity, metabolic reprogramming, and m6A methylation, the utilization of data remains insufficient. Furthermore, the uniqueness and limitations of the selected modeling methods, coupled with the lack of rigorous validation in large multicenter cohorts, have hindered their widespread application in clinical settings.

TLRs play a crucial role as key innate immune sensors in identifying PAMPs and damage-associated molecular patterns (DAMPs) (34). TLRs regulate tumorigenesis by activating various immune cells (35). Recently, there has been increasing attention on the role and mechanisms of TLRs in the progression of specific cancers (36, 37), which has led to the development of various therapeutic strategies targeting TLRs. Several clinical trials have demonstrated that combination immunotherapy involving TLR agonists significantly enhances the treatment response in cancer patients (38, 39).

This study innovatively integrates mRNA expression profiles, DNA methylation data, and somatic mutation information related to the Toll-like receptor pathway, successfully establishing a comprehensive consensus subtype classification system for lung adenocarcinoma. Further analysis of subtype-specific differential expression was conducted, and an AIDPI scoring system was constructed using 101 machine learning algorithms. Clinical prognostic analysis indicates that high AIDPI scores are significantly correlated with shortened patient survival and advanced tumor progression. Additionally, we found that AIDPI is associated with the mutation status of classic tumor driver genes such as TP53 and EGFR. Previous studies have confirmed that mutations in these genes can enhance the sensitivity of lung adenocarcinoma to immunotherapy (40, 41). These research

results not only enrich the molecular subtyping system of lung adenocarcinoma but also provide a scientific basis for optimizing precise stratified diagnosis and treatment strategies.

In addition to introducing AIDPI for the stratification of high-risk LUAD patients, we also identified NPC2 as a potential therapeutic target for this patient population. In recent years, multiple studies have revealed the significant role of abnormal cholesterol metabolism in the occurrence and development of tumors. The imbalance in cholesterol homeostasis is closely associated with the increased survival rates and inhibited apoptosis of malignant tumor cells (42), and it also significantly promotes tumor formation (43). In the cholesterol transport mechanism, NPC2 acts as a key protein that regulates the excretion and transport of cholesterol from lysosomes through its interaction with membrane proteins (44–46). Research has demonstrated that the activation of SREBP-1 can induce autophagy and increase NPC2 expression in cholesterol-depleted states, thereby facilitating lysosomal cholesterol release and accelerating glioblastoma progression (47). Notably, NPC2 effectively blocks the infiltration of stromal macrophage lines into early lung cancer tissues by regulating intracellular cholesterol balance and inhibiting CCL6 secretion (48). During liver cancer development, abnormal NPC2 expression regulates the proliferation, migration, and tumor formation of cancer cells by influencing the activation of the ERK1/2 signaling pathway (49). Beyond its role in tumor progression, NPC2 is also implicated in mechanisms of chemotherapy resistance. Research indicates that the downregulation of NPC2 expression leads to the accumulation of free cholesterol, which enhances the MAPK/AKT signaling pathway in hepatocellular carcinoma, thereby reducing the therapeutic efficacy of sorafenib (50). Furthermore, ALKBH5 plays a regulatory role in oxaliplatin resistance in colorectal cancer by inhibiting the YTHDF2-mediated m6A demethylation of NPC2 (51). However, the biological functions and associated molecular regulatory mechanisms of NPC2 in LUAD remain poorly understood. In this study, we demonstrated that elevated NPC2 expression correlates with favorable prognosis in the LUAD cohort. Furthermore, we discovered that the overexpression of NPC2 in LUAD cell lines results in decreased proliferation and enhanced apoptosis, which is attributed to the inhibition of the PI3K/Akt signaling pathway.

Cyclin-dependent kinases (CDKs) are essential for the regulation of the cell cycle (52). In recent years, Ribociclib, a targeted inhibitor of CDK4/6, has been developed and clinically evaluated. This therapy has notably extended progression-free survival and overall survival for patients battling advanced breast cancer (53, 54). However, despite the improved prognosis associated with CDK4/6 inhibitor treatment, approximately 10% of tumors exhibit initial resistance when used as monotherapy, resulting in a poor prognosis (55–57). Previous research has shown that combining Ribociclib with Alpelisib (PI3K inhibitor) yields a favorable synergistic effect in early breast cancer and colorectal cancer models (58, 59). Notably, our research suggests that NPC2, as a novel biomarker, can inhibit the growth of lung adenocarcinoma by suppressing the PI3K signaling pathway and

enhance the sensitivity of Ribociclib in the treatment of lung adenocarcinoma. Furthermore, our single-cell analysis also found that NPC2 is mainly expressed in monocytes/macrophages, epithelial cells, and fibroblasts in the LUAD microenvironment. Previous studies have shown that based on ICD markers (60) and exhausted CD8⁺ T cell features (61), further integrated scRNA seq data to construct prognostic features through machine learning, both enhancing clinical prediction. Complementarily, spatial transcriptomics has revealed M2a-type TAMs as drivers of spread through air spaces via the CCL17-CCR4/ β -catenin axis (62), suggesting precise therapeutic strategies for STAS-positive patients. Collectively, these studies contribute to our understanding of LUAD immunobiology and patient stratification.

Compared to earlier published studies, our research presents several significant differences. This study integrates omics information from all three dimensions of lung adenocarcinoma for the first time and employs ten clustering algorithms for a comprehensive analysis, effectively mitigating the influence of single algorithm preferences on the research findings. Additionally, the modeling factors derived from multi-queue prognostic gene screening considerably enhance the robustness and predictive performance of the model. Furthermore, by integrating data from four multicenter lung adenocarcinoma cohorts and combining them with ten machine learning algorithms, we constructed an AIDPI model aimed at minimizing the risk of overfitting. However, this study does have certain limitations. Firstly, despite employing calibration algorithms for data standardization, heterogeneity in sample size and sequencing platforms persists within the queue, affecting model performance. Therefore, further validation is required in a prospective, multicenter, and technically standardized cohort. Secondly, all samples were obtained from retrospective studies, and comprehensive validation of the AIDPI model's clinical applicability is necessary through extensive prospective multicenter research. Lastly, the absence of certain clinical and molecular feature information in public datasets may impact the potential correlation analysis between AIDPI and related variables.

5 Conclusion

In summary, this study successfully established a prognostic evaluation model for lung adenocarcinoma by integrating Toll-like receptor signaling-related genes, employing bioinformatics analysis and machine learning algorithms. This research establishes a scientific basis for the creation of customized treatment strategies for LUAD patients with varying risk levels in clinical practice. Furthermore, we elucidated the carcinogenic mechanism of NPC2 in lung adenocarcinoma, confirming its potential as an effective biological indicator for the prognosis of LUAD patients. However, the specific mechanism of NPC2 in the process of LUAD metastasis requires further exploration using metastatic nude mouse models in future studies.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#). Further inquiries can be directed to the corresponding authors.

Ethics statement

The animal study was approved by the Animal Ethics and Welfare Committee of North China University of Science and Technology (2024-SY-003). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

AL: Conceptualization, Validation, Writing – original draft. PC: Funding acquisition, Writing – review & editing, Data curation. LL: Validation, Methodology, Writing – original draft. JL: Validation, Writing – original draft, Methodology. XZ: Methodology, Validation, Writing – original draft. WW: Data curation, Software, Writing – original draft. ZY: Writing – original draft, Software, Data curation. YG: Writing – review & editing, Funding acquisition, Project administration, Supervision. HZ: Writing – review & editing, Project administration, Funding acquisition, Supervision.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was supported by Hebei Provincial Department of Science and Technology Centrally Guided Local Development Fund Project (246Z7712G), the Applied and Basic Research Program from Tangshan Science and Technology Bureau (24130222C), the Basic Scientific Research Funds for Provincial Universities of North China University of Science and Technology (JJC2024053) and the Medical and Health Science and Technology Development Project of Shandong (2019WS353).

Conflict of interest

The authors declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1697560/full#supplementary-material>

References

- Testa U, Castelli G, Pelosi E. Lung cancers: molecular characterization, clonal heterogeneity and evolution, and cancer stem cells. *Cancers (Basel)*. (2018) 10(8). doi: 10.3390/cancers10080248
- Salto A, Shafique M, Chiappori A. Update on the biology, management, and treatment of small cell lung cancer (SCLC). *Front Oncol*. (2020) 10:1074. doi: 10.3389/fonc.2020.01074
- Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. (2012) 150:1107–20. doi: 10.1016/j.cell.2012.08.029
- Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol*. (2016) 17:e542–51. doi: 10.1016/S1470-2045(16)30406-5
- Chan TA, Yarchoon M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol*. (2019) 30:44–56. doi: 10.1093/annonc/mdy495
- Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun*. (2017) 8:15180. doi: 10.1038/ncomms15180
- Karin M, Shalpour S. Regulation of antitumor immunity by inflammation-induced epigenetic alterations. *Cell Mol Immunol*. (2022) 19:59–66. doi: 10.1038/s41423-021-00756-y
- Vezzani B, Carinci M, Prevati M, Giacomazzi S, Della Sala M, Gafà R, et al. Epigenetic regulation: A link between inflammation and carcinogenesis. *Cancers (Basel)*. (2022) 14(5). doi: 10.3390/cancers14051221
- Candido S, Tomasello BMR, Lavoro A, Falzone L, Gattuso G, Libra M, et al. Novel insights into epigenetic regulation of IL6 pathway: in silico perspective on inflammation and cancer relationship. *Int J Mol Sci*. (2021) 22(18). doi: 10.3390/ijms221810172
- Hajishengallis G, Lambris JD. Microbial manipulation of receptor crosstalk in innate immunity. *Nat Rev Immunol*. (2011) 11:187–200. doi: 10.1038/nri2918
- Fitzgerald KA, Kagan JC. Toll-like receptors and the control of immunity. *Cell*. (2020) 180:1044–66. doi: 10.1016/j.cell.2020.02.041
- Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol*. (2010) 11:373–84. doi: 10.1038/ni.1863
- Gauzzi MC, Del Cornò M, Gessani S. Dissecting TLR3 signalling in dendritic cells. *Immunobiology*. (2010) 215:713–23. doi: 10.1016/j.imbio.2010.05.008
- Schreibelt G, Tel J, Sliepen KH, Benitez-Ribas D, Figdor CG, Adema GJ, et al. Toll-like receptor expression and function in human dendritic cell subsets: implications for dendritic cell-based anti-cancer immunotherapy. *Cancer Immunol Immunother*. (2010) 59:1573–82. doi: 10.1007/s00262-010-0833-1
- Kim HS, Lee JH, Nam SJ, Ock CY, Moon JW, Yoo CW, et al. Association of PD-L1 expression with tumor-infiltrating immune cells and mutation burden in high-grade neuroendocrine carcinoma of the lung. *J Thorac Oncol*. (2018) 13:636–48. doi: 10.1016/j.jtho.2018.01.008
- Yi M, Jiao D, Xu H, Liu Q, Zhao W, Han X, et al. Biomarkers for predicting efficacy of PD-1/PD-L1 inhibitors. *Mol Cancer*. (2018) 17:129. doi: 10.1186/s12943-018-0864-3
- Urban-Wojcik Z, Khan MM, Oyler BL, Fähræus R, Marek-Trzonkowska N, Nita-Lazar A, et al. The role of TLRs in anti-cancer immunity and tumor rejection. *Front Immunol*. (2019) 10:2388. doi: 10.3389/fimmu.2019.02388
- Bertoglio P, Aprile V, Ventura L, Cattoni M, Nachira D, Lococo F, et al. Impact of high-grade patterns in early-stage lung adenocarcinoma: A multicentric analysis. *Lung*. (2022) 200:649–60. doi: 10.1007/s00408-022-00561-y
- Kadota K, Kushida Y, Kagawa S, Ishikawa R, Ibuki E, Inoue K, et al. Cribriform subtype is an independent predictor of recurrence and survival after adjustment for the eighth edition of TNM staging system in patients with resected lung adenocarcinoma. *J Thorac Oncol*. (2019) 14:245–54. doi: 10.1016/j.jtho.2018.09.028
- Koncina E, Haan S, Rauh S, Letellier E. Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. *Cancers (Basel)*. (2020) 12(2). doi: 10.3390/cancers12020319
- Chu G, Ji X, Wang Y, Niu H. Integrated multiomics analysis and machine learning refine molecular subtypes and prognosis for muscle-invasive urothelial cancer. *Mol Ther Nucleic Acids*. (2023) 33:110–26. doi: 10.1016/j.omtn.2023.06.001
- Xie Y, Chen H, Tian M, Wang Z, Wang L, Zhang J, et al. Integrating multi-omics and machine learning survival frameworks to build a prognostic model based on immune function and cell death patterns in a lung adenocarcinoma cohort. *Front Immunol*. (2024) 15:1460547. doi: 10.3389/fimmu.2024.1460547
- Meng J, Jiang A, Lu X, Gu D, Ge Q, Bai S, et al. Multiomics characterization and verification of clear cell renal cell carcinoma molecular subtypes to guide precise chemotherapy and immunotherapy. *Imeta*. (2023) 2:e147. doi: 10.1002/imt2.147
- Storch J, Xu Z. Niemann-Pick C2 (NPC2) and intracellular cholesterol trafficking. *Biochim Biophys Acta*. (2009) 1791:671–8. doi: 10.1016/j.bbalip.2009.02.001
- Liu X, Jia Y, Shi C, Kong D, Wu Y, Zhang T, et al. CYP4B1 is a prognostic biomarker and potential therapeutic target in lung adenocarcinoma. *PloS One*. (2021) 16:e0247020. doi: 10.1371/journal.pone.0247020
- Li Q, Liu XL, Jiang N, Li QY, Song YX, Ke XX, et al. A new prognostic model for RHOV, ABCC2, and CYP4B1 to predict the prognosis and association with immune infiltration of lung adenocarcinoma. *J Thorac Dis*. (2023) 15:1919–34. doi: 10.21037/jtd-23-265
- Peng Z, Wang J, Shan B, Li B, Peng W, Dong Y, et al. The long noncoding RNA LINC00312 induces lung adenocarcinoma migration and vasculogenic mimicry through directly binding YBX1. *Mol Cancer*. (2018) 17:167. doi: 10.1186/s12943-018-0920-z
- Li J, Wang J, Chen Y, Li S, Jin M, Wang H, et al. LncRNA MALAT1 exerts oncogenic functions in lung adenocarcinoma by targeting miR-204. *Am J Cancer Res*. (2016) 6:1099–107.
- Oh M, Park S, Kim S, Chae H. Machine learning-based analysis of multi-omics data on the cloud for investigating gene regulations. *Brief Bioinform*. (2021) 22:66–76. doi: 10.1093/bib/bbaa032
- Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol Adv*. (2021) 49:107739. doi: 10.1016/j.biotechadv.2021.107739
- Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun*. (2022) 13:816. doi: 10.1038/s41467-022-28421-6
- Chen S, Jiang L, Zheng X, Shao J, Wang T, Zhang E, et al. Clinical use of machine learning-based pathomics signature for diagnosis and survival prediction of bladder cancer. *Cancer Sci*. (2021) 112:2905–14. doi: 10.1111/cas.14927
- Han X, Guo Y, Ye H, Chen Z, Hu Q, Wei X, et al. Development of a machine learning-based radiomics signature for estimating breast cancer TME phenotypes and predicting anti-PD-1/PD-L1 immunotherapy response. *Breast Cancer Res*. (2024) 26:18. doi: 10.1186/s13058-024-01776-y
- Pradere JP, Dapito DH, Schwabe RF. The Yin and Yang of Toll-like receptors in cancer. *Oncogene*. (2014) 33:3485–95. doi: 10.1038/onc.2013.302
- Cao LL, Kagan JC. Targeting innate immune pathways for cancer immunotherapy. *Immunology*. (2023) 56:2206–17. doi: 10.1016/j.immuni.2023.07.018

36. Rakoff-Nahoum S, Medzhitov R. Toll-like receptors and cancer. *Nat Rev Cancer*. (2009) 9:57–63. doi: 10.1038/nrc2541
37. Pahlavanneshan S, Sayadmanesh A, Ebrahimiyan H, Basiri M. Toll-like receptor-based strategies for cancer immunotherapy. *J Immunol Res*. (2021) 2021:9912188. doi: 10.1155/2021/9912188
38. Koh J, Kim S, Lee SN, Kim SY, Kim JE, Lee KY, et al. Therapeutic efficacy of cancer vaccine adjuvanted with nanoemulsion loaded with TLR7/8 agonist in lung cancer model. *Nanomedicine*. (2021) 37:102415. doi: 10.1016/j.nano.2021.102415
39. Ribas A, Medina T, Kirkwood JM, Zakharia Y, Gonzalez R, Davar D, et al. Overcoming PD-1 blockade resistance with cpG-A toll-like receptor 9 agonist vidutolimod in patients with metastatic melanoma. *Cancer Discov*. (2021) 11:2998–3007. doi: 10.1158/2159-8290.Cd-21-0425
40. Zhang C, Yang C, Shi Q. Effects of tp53 gene mutations on the survival of non-small cell lung cancer (NSCLC); A short review. *Cancer Manag Res*. (2025) 17:65–82. doi: 10.2147/cmar.S495006
41. Lui K, Cheung KK, Ng WW, Wang Y, Au DWH, Cho WC, et al. The impact of genetic mutations on the efficacy of immunotherapies in lung cancer. *Int J Mol Sci*. (2024) 25(2). doi: 10.3390/ijms252211954
42. Yue S, Li J, Lee SY, Lee HJ, Shao T, Song B, et al. Cholesteryl ester accumulation induced by PTEN loss and PI3K/AKT activation underlies human prostate cancer aggressiveness. *Cell Metab*. (2014) 19:393–406. doi: 10.1016/j.cmet.2014.01.019
43. Llaverrias G, Danilo C, Mercier I, Daumer K, Capozza F, Williams TM, et al. Role of cholesterol in the development and progression of breast cancer. *Am J Pathol*. (2011) 178:402–12. doi: 10.1016/j.ajpath.2010.11.005
44. Liao Y, Wei J, Wang J, Shi X, Luo J, Song BL, et al. The non-canonical NF- κ B pathway promotes NPC2 expression and regulates intracellular cholesterol trafficking. *Sci China Life Sci*. (2018) 61:1222–32. doi: 10.1007/s11427-018-9339-0
45. Qian H, Wu X, Du X, Yao X, Zhao X, Lee J, et al. Structural basis of low-pH-dependent lysosomal cholesterol egress by NPC1 and NPC2. *Cell*. (2020) 182:98–111.e18. doi: 10.1016/j.cell.2020.05.020
46. Xu Y, Zhang Q, Tan L, Xie X, Zhao Y. The characteristics and biological significance of NPC2: Mutation and disease. *Mutat Res Rev Mutat Res*. (2019) 782:108284. doi: 10.1016/j.mrrev.2019.108284
47. Geng F, Guo D. SREBF1/SREBP-1 concurrently regulates lipid synthesis and lipophagy to maintain lipid homeostasis and tumor growth. *Autophagy*. (2024) 20:1183–5. doi: 10.1080/15548627.2023.2275501
48. Kamata T, Jin H, Giblett S, Patel B, Patel F, Foster C, et al. The cholesterol-binding protein NPC2 restrains recruitment of stromal macrophage-lineage cells to early-stage lung tumours. *EMBO Mol Med*. (2015) 7:1119–37. doi: 10.15252/emmm.201404838
49. Liao YJ, Fang CC, Yen CH, Hsu SM, Wang CK, Huang SF, et al. Niemann-Pick type C2 protein regulates liver cancer progression via modulating ERK1/2 pathway: Clinicopathological correlations and therapeutic implications. *Int J Cancer*. (2015) 137:1341–51. doi: 10.1002/ijc.29507
50. Suk FM, Wang YH, Chiu WC, Liu CF, Wu CY, Chen TL, et al. Secretory NPC2 protein-mediated free cholesterol levels were correlated with the sorafenib response in hepatocellular carcinoma. *Int J Mol Sci*. (2021) 22(16). doi: 10.3390/ijms22168567
51. Wan P, Ren Y, Li HT. ALKBH5-mediated NPC2 mRNA m(6)A demethylation promotes resistance to oxaliplatin in colorectal cancer. *Funct Integr Genomics*. (2025) 25:158. doi: 10.1007/s10142-025-01651-9
52. Icard P, Fournel L, Wu Z, Alifano M, Lincet H. Interconnection between metabolism and cell cycle in cancer. *Trends Biochem Sci*. (2019) 44:490–501. doi: 10.1016/j.tibs.2018.12.007
53. Rampioni Vinciguerra GL, Sonogo M, Segatto I, Dall'Acqua A, Vecchione A, Baldassarre G, et al. CDK4/6 inhibitors in combination therapies: better in company than alone: A mini review. *Front Oncol*. (2022) 12:891580. doi: 10.3389/fonc.2022.891580
54. Piezzo M, Cocco S, Caputo R, Cianniello D, Gioia GD, Lauro VD, et al. Targeting cell cycle in breast cancer: CDK4/6 inhibitors. *Int J Mol Sci*. (2020) 21(18). doi: 10.3390/ijms21186479
55. Krasniqi E, Goeman F, Pulito C, Palcau AC, Ciuffreda L, Di Lisa FS, et al. Biomarkers of response and resistance to CDK4/6 inhibitors in breast cancer: hints from liquid biopsy and microRNA exploration. *Int J Mol Sci*. (2022) 23(23). doi: 10.3390/ijms232314534
56. Michaloglou C, Crafter C, Siersbaek R, Delpuech O, Curwen JO, Carnevalli LS, et al. Combined inhibition of mTOR and CDK4/6 is required for optimal blockade of E2F function and long-term growth inhibition in estrogen receptor-positive breast cancer. *Mol Cancer Ther*. (2018) 17:908–20. doi: 10.1158/1535-7163.Mct-17-0537
57. Fritsch C, Huang A, Chatenay-Rivauday C, Schnell C, Reddy A, Liu M, et al. Characterization of the novel and specific PI3K α inhibitor NVP-BYL719 and development of the patient stratification strategy for clinical trials. *Mol Cancer Ther*. (2014) 13:1117–29. doi: 10.1158/1535-7163.Mct-13-0865
58. Aslam R, Richards CE, Fay J, Hudson L, Workman J, Lee CL, et al. Synergistic effects of the combination of alpelisib (PI3K inhibitor) and ribociclib (CDK4/6 inhibitor) in preclinical colorectal cancer models. *Int J Mol Sci*. (2024) 25(24). doi: 10.3390/ijms252413264
59. Tolaney SM, Im YH, Calvo E, Lu YS, Hamilton E, Forero-Torres A, et al. Phase Ib Study of Ribociclib plus Fulvestrant and Ribociclib plus Fulvestrant plus PI3K Inhibitor (Alpelisib or Buparlisib) for HR(+) Advanced Breast Cancer. *Clin Cancer Res*. (2021) 27:418–28. doi: 10.1158/1078-0432.Ccr-20-0645
60. Zhang H, Mu Q, Jiang Y, Zhao X, Jia X, Wang K, et al. Integrative single-cell and machine learning approach to characterize immunogenic cell death and tumor microenvironment in LUAD. *J Transl Med*. (2025) 23:1000. doi: 10.1186/s12967-025-06889-2
61. Zhang H, Zhang P, Lin X, Tan L, Wang Y, Jia X, et al. Integrative single-cell analysis of LUAD: elucidating immune cell dynamics and prognostic modeling based on exhausted CD8 $^{+}$ T cells. *Front Immunol*. (2024) 15:1366096. doi: 10.3389/fimmu.2024.1366096
62. Ding Y, Zhang H, Wen H, Zhao S, Li J, Liu X, et al. Single-cell and spatial transcriptomics-based research reveals association between M2a macrophages and tumor spread through air spaces in lung adenocarcinoma. *Am J Pathol*. (2025) 195:2477–91. doi: 10.1016/j.ajpath.2025.07.017