scientific reports

Check for updates

OPEN

Integrated machine learning survival framework for consensus modeling in a large multicenter cohort of NSCLC resistant to aumolertinib

Xiao Wu[1], Yang Lu[2], Yongping Li[3], Annan Zhu[4], Tingting Wang[1], Fang Su[1], Zishu Wang[1], Yongxia Chen[1] & Guolei Song[1]✉

Patients with advanced non-small cell lung cancer (NSCLC) harboring epidermal growth factor receptor (EGFR) mutations often benefit from third-generation tyrosine kinase inhibitors (TKIs), such as aumolertinib (AUM). However, the development of drug resistance significantly limits the clinical efficacy of AUM. To address this, we established an in vitro model of AUM-resistant cell lines and performed RNA sequencing to identify resistance-associated differentially expressed genes. Using machine learning, we constructed an AUM resistance-related prognostic signature (ARRPS). Our results demonstrated that ARRPS effectively predicts the prognostic risk of patients. Notably, for patients with high ARRPS scores, the addition of CD-437 or TPCA-1 to conventional AUM treatment may help overcome drug resistance. These findings suggest that ARRPS serves as both a prognostic tool and a guide for personalized treatment strategies, potentially optimizing the clinical management of NSCLC patients.

Keywords NSCLC, Resistance, Prognostic signature, Aumolertinib

Abbreviations

NSCLC	Non-small cell lung cancer
EGFR-TKI	Epidermal growth factor receptor-tyrosine kinase inhibitor
AUM	Aumolertinib
ARRPS	Aumolertinib resistance-related prognostic signature
LUAD	Lung adenocarcinoma

The problem of drug resistance to cancer treatment has long been recognized as a key factor in cancer treatment failure[1]. For example, resistance to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) greatly contributes to treatment failure and cancer recurrence in patients with EGFR-mutant non-small cell lung cancer (NSCLC)[2]. Among the multiple mechanisms of resistance to EGFR-TKIs, T790M, a secondary mutation in EGFR exon 20, is thought to be the most prevalent cause of acquired resistance to first- and second-generation EGFR-TKIs, and a number of third-generation EGFR TKIs have been designed to irreversibly target EGFRs with both T790M-resistant mutations and EGFR-activating mutations. Among them, aumolertinib (AUM) is a representative compound widely approved for standard second-line treatment of EGFR T790M-mutant NSCLC and first-line treatment of EGFR-activating mutant NSCLC, with significant preclinical and clinical efficacy but inevitably acquired resistance. Multiple mechanisms have been reported, including EGFR mutations[2]activation of alternative pathways[3]and aberrant downstream signaling[4]. However, the investigation of AUM resistance mechanisms remains an urgent problem, which simultaneously attracts people to explore new resistance mechanisms and find potential therapeutic strategies for AUM-resistant NSCLC.

[1]The First Affiliated Hospital of Bengbu Medical University, Bengbu, China. [2]College of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan, China. [3]School of Pharmacy, Bengbu Medical University, Bengbu, China. [4]Computer Major, Northeastern University, Qinhuangdao, China. ✉email: songguolei2008@163.com

In order to be able to better find the mechanisms of resistance to AUM as well as to have an accurate prediction of the prognosis of NSCLC patients, we developed and validated a robust signature consisting of a small set of genes using large-scale data. Using 10 machine learning algorithms, they were transformed into 100 combinations that were further executed in 5 independent cohorts. The best-performing ARRPS among the 100 model types was finally identified. In this model, patients with a high ARRPS showed a higher risk of death, relapse, and disease progression[5]. The ARRPS was independent of traditional clinical signatures (e.g., American Joint Committee on Cancer (AJCC) staging) and molecular signatures (e.g., TP53 mutations) and additionally demonstrated superior predictive performance in these variables. Comparisons with published signatures show favorable performance of ARRPS. Additionally, patients with high ARRPS scores may benefit from 2 potential therapeutic agents: CD-437 and TPCA-1.

Methods

Induction of AUM-resistant NSCLC cell lines

At the beginning, we modeled acquired resistance to AUM by deriving polyclonal acquired resistant cell lines on the basis of stepwise dose escalation over a period of six months followed by maintenance in 10μM of drug over 6 weeks[6] that is well representative of the clinic, and performed RNA whole transcriptome sequencing of this model using parental cells as a control to screen for AUM-resistant genes.

Generation of ARRPS

We extracted data from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) (GSE50081, GSE30219, GSE72094 and GSE132313) and The Cancer Genome Atlas (TCGA, http://portal.gdc.cancer.gov/) (TCGA-LUAD) in which five independent lung adenocarcinoma datasets were retrospectively extracted. Sample selection criteria were as follows (1) primary cancer tissues that had not received any adjuvant therapy and (2) survival information and RNA expression data were available. In the TCGA-LUAD project, we obtained single nucleotide variants (SNV) and copy number variants (CNV) from the TCGA portal. Prognostic genes were screened using Cox regression analysis, and genes with an adjusted P-value < 0.05 and a consistent hazard ratio (HR) direction in the four cohorts were considered to be robust genes associated with overall survival (OS). The intersection of differentially expressed genes in AUM-resistant cell lines with the above obtained genes was used to identify robust genes associated with AUM resistance and OS. To develop a consensus ARRPS with high accuracy and stable performance, we integrated 10 machine learning algorithms including Random Survival Forest (RSF), Elastic Network (Enet), Lasso regression, Ridge regression, Stepwise Cox regression, Cox boosting, Cox partial least squares regression (plsRcox), Supervised Principal Component Analysis (SuperPC), Generalized Augmented Regression Modeling (GBM) and Survival Support Vector Machine (survival-svm). In this study, the TCGA-LUAD cohort was used as the training set, while the other four cohorts were used as the test set. We calculated the c-index of each feature in all cohorts, where the feature with the highest average c-index was considered as the best feature[7].

Validating the prognostic value of ARRPS

Patients in the five cohorts were categorized into high and low ARRPS subgroups based on the median ARRPS score. Kaplan-Meier curves and Cox regression analysis were used to assess the prognostic value of ARRPS. Subjects' work characteristics (ROC) curves were plotted to assess the predictive accuracy of ARRPS. Subsequently, we calculated the C-index to compare the superiority of ARRPS with a variety of common clinical characteristics (including gender, age, TNM stage and grading) in predicting overall survival (OS) in patients with LUAD.

Collection and calculation of LUAD published signatures

To compare the predictive performance of ARRPS with these published features, we searched PubMed on October 1, 2024, for published articles on prognostic models[8–11]. Subsequently, we calculated risk scores for the 5 cohorts based on the genes and coefficients provided in the articles and assessed their prognostic performance in LUAD by one-way Cox analysis and C-index.

Genomic alteration landscape

To characterize somatic variations underlying ARRPS stratification, we analyzed genomic alterations in the TCGA-LUAD cohort. Single nucleotide variants (SNVs) were processed using GATK4 and annotated via ANNOVAR. Copy number variations (CNVs) were called using GISTIC2.0. Differential mutation frequencies between high/low ARRPS groups were identified through Fisher's exact tests (FDR < 0.05), while CNV segment disparities were assessed via Wilcoxon rank-sum tests. Significant loci were visualized using maftools and Integrative Genomics Viewer. Co-occurring SNV pairs were detected using DISCOVER, with permutation testing ($p < 0.01$). All analyses were conducted in R/Bioconductor and Genepattern (genepattern.org).

Gene set variation analysis (GSVA)

Gene sets were obtained from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). Gene set variance analysis (GSVA) methods were used to quantify the values of each metric in each sample. The limma package was used to identify pathways that were significantly altered between the high and low ARRPS groups[12].

Comprehensive analyses based on immune cell infiltration

We explored the correlation between ARRPS scores and immune infiltration. We obtained immune infiltration data for TCGA-LUAD from the TIMER 2.0 web server (http://timer.cistrome.org/)[13], which covers immune

infiltration data obtained by seven methods, namely, CIBERSORT, CIBERSORT-ABS, EPIC, MCPCOUNTER, QUANTISEQ, TIMER and XCELL[14–17].

Development and validation of potential therapeutic agents

We obtained drug sensitivity data for ARRPS from the CTRP and PRISM recombinant datasets. Based on the Wilcoxon rank-sum test, we analyzed the difference in drug response between the high ARRPS group (top 10%) and the low ARRPS group (bottom 10%) and set a threshold of log2FC < 0.05 to identify compounds with lower AUC values in the high ARRPS group. Next, we applied Spearman's correlation analysis to further screen compounds whose AUC values were negatively correlated with ARRPS (setting a threshold $R < -0.3$)[18–21].

Tumor xenograft models

All animals were housed according to the guidelines of the relevant institutions and approved by the Animal Care Committee of Bengbu Medical College. The BALB/c nude mice (male, 6 weeks old) were purchased from Unilever (Beijing, China). The BALB/c nude mice were kept in a specific pathogen-free (SPF)-grade animal laboratory. Nude mice were injected subcutaneously with 5×10^6 cells on the right side and randomly grouped after the tumors grew to 100–150 cubic millimeters. Tumor volume was calculated as length × width^2 × 0.5. After three weeks, all mice were executed by carbon dioxide asphyxiation, and tumors were harvested and weighed.

Cell lines and culture conditions

The HCC827 and H1975 cell line was sourced from American Type Culture Collection (Manassas, VA, USA). All media were supplemented with 10% fetal bovine serum (FBS; Gibco), 100 U/ml penicillin, and 100 μg/ml streptomycin. The cells were incubated at 37 °C in a humidified 5% CO2 atmosphere. The authenticity of the human cell lines used in this study was ascertained through short tandem repeat (STR) profiling.

RNA sequencing

RNA Sequencing and data collection were performed by Wei Huan Biotechnology (China) Co. Total RNA was extracted using TRIzol® reagent (Invitrogen, #15596018) according to the manufacturer's instructions, and the samples were sent to Shanghai Wei Huan Biotechnology (Shanghai, China) for RNA purification, reverse transcription, library construction, and sequencing. After quality control of raw reads, clean reads were mapped using HISAT2 software and further assembled using StringTiein. Cellular pathway analysis was performed using GSEA.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 8, and data are expressed as mean ± SEM. Two-tailed unpaired Student's t-test and one- or two-way ANOVA were used to analyze differences between groups.

Results

Construction of the ARRPS

The ARRPS construction process is shown in Fig. 1. HCC827 cells resistant to AUM were obtained through six months of induction with a resistance index of 3.35, which is a good representation of the clinical situation (Fig. 2A). We performed RNA whole transcriptome sequencing of this model using parental cells as a control to screen for AUM resistance genes. We found 2987 up-regulated genes and 3410 down-regulated genes in resistant HCC827 cells (Fig. 2B-C). Through analysis, we obtained 20 genes significantly associated with overall survival and aumolertinib resistance, including ANLN, ASPM, BUB1B, CDC20, CDC25C, CDC6, CDCA2, CDCA3, CDKN3, CENPE, CENPF, FOXM1, GINS4, KIF2C, MELK, MYBL2, NEK2, PRC1, SHCBP1, and TOP2A.Based on the 10-fold cross-validation framework, we used 10 machine learning algorithms and their constituent combinations for a total of 100 methods for modeling. The results show that the combined algorithm of lasso + RSF (including 12 genes) has the highest average C-index and is considered the optimal modeling approach (Fig. 2D). We found that these 12 genes were highly expressed in cancer tissues (Fig. 2E), and that these 12 genes had a high ability to discriminate between diagnosing LUAD and non-tumors (Fig. 2F). To evaluate the prognostic performance of ARRPS, we categorized LUAD patients into high and low ARRPS groups according to the median value. KM curves showed that in the training cohort, mortality was significantly higher in the high-risk group than in the low-risk group (TCGA-LUAD, Fig. 2G), and the other four validation cohorts showed the same trend of patients in the high-ARRPS group having worse OS (Fig. 2H-K). For the five cohorts, the ROC curves demonstrated the high ability of ARRPS to predict patient OS (Fig. 3A-E). Subsequently, the ability of ARRPS to predict OS in patients with LUAD was compared with a variety of clinical characteristics, showing that ARRPS was significantly more accurate than these characteristics, including age, gender, smoking BI, TNM staging, KRAS, EGFR, STK11, and TP53 mutations or de novo events (Fig. 3F-J), and the results were tabulated (Table 1).Meanwhile, we also investigated other clinical outcomes of TCGA-LUAD, including DSS, PFI, and DFI, and the results showed that patients with high ARRPS had worse DSS, PFI, and DFI, and ARRPS was an independent risk factor for DSS, PFI, and DFI in TCGA-LUAD. ARRPS also had high accuracy in predicting TCGA-LUAD DSS, PFI, and DFI (Fig. 4A-J). Critically, univariate and multivariate Cox regression analyses confirmed ARRPS as an independent risk factor for PFI (Fig. 4K-L), reinforcing its prognostic robustness across multiple endpoints. In multiple cohorts, as described above, our ARRPS model shows robust pre-prediction performance. To further verify whether ARRPS has good predictive performance compared to published signatures, we collected 40 published signatures. We then calculated the c-index for each of the 40 signatures in each of the five cohorts and compared them to ARRPS. The results show that the c-index of ARRPS is significantly higher than that of other signatures in the TCGA-LUAD, GSE30219, GSE50081, and GSE72094 cohorts, except for the GSE13213 cohort (Fig. S1A). This suggests that ARRPS maintains a relatively strong

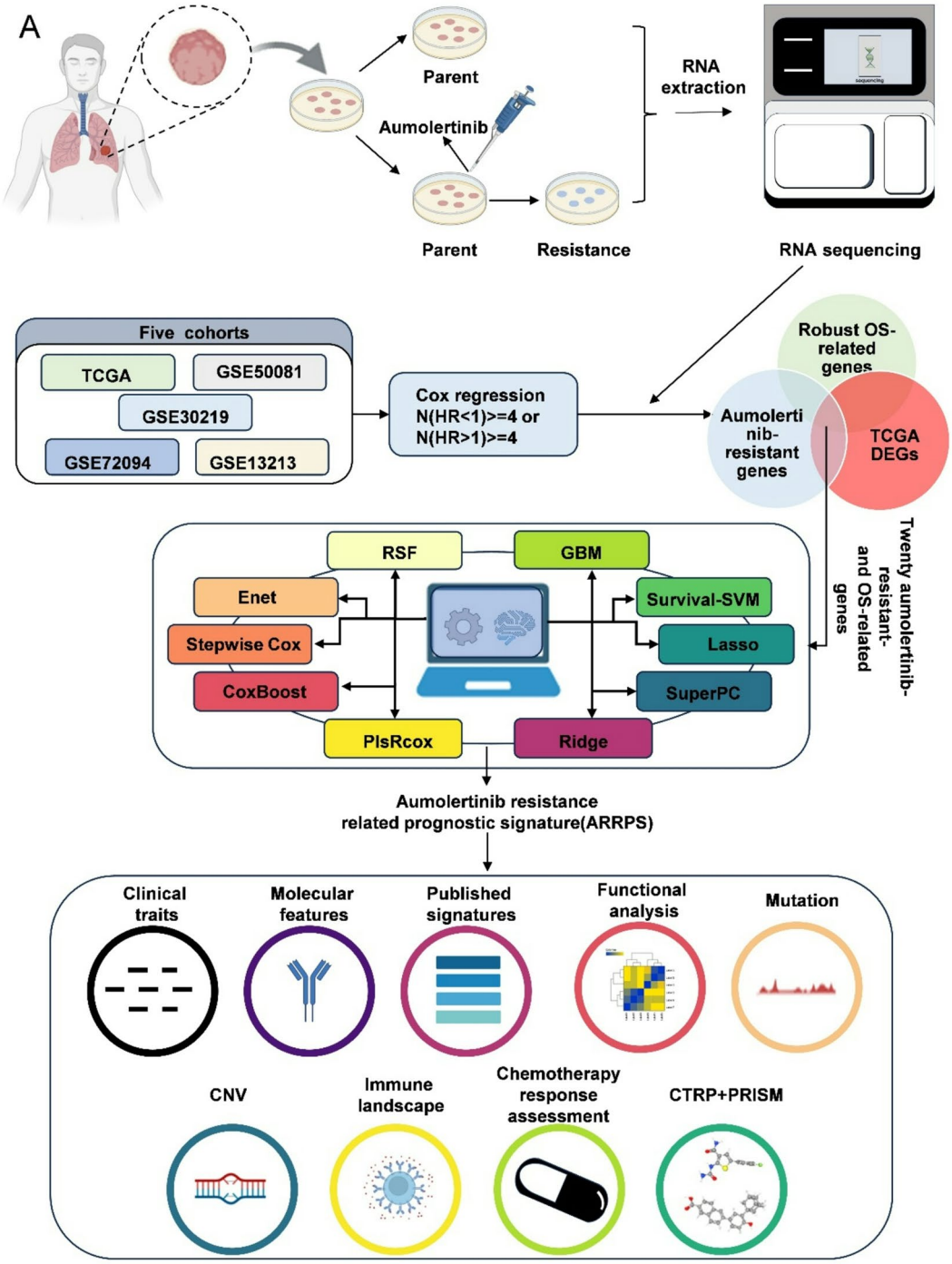

Fig. 1. The flowchart for building ARRPS.

predictive power even when compared to other signatures. Taken together, the results of Kaplan-Meier survival analyses, Cox regression analyses, ROC curves, and C-indexes of the five cohorts consistently showed that the ARRPS accurately and robustly predicted the prognosis of patients with LUAD, suggesting that the ARRPS may become an attractive tool to serve clinical practice.

ARRPS shows substantial correlation with immune-relate

To explore the immune status reflected by the ARRPS profile score, we analyzed the differences between high and low ARRPS profile scores in immune-infiltrating cells, which were concentrated in T cells, B cells, macrophages, and NK cells (Fig. 5A).

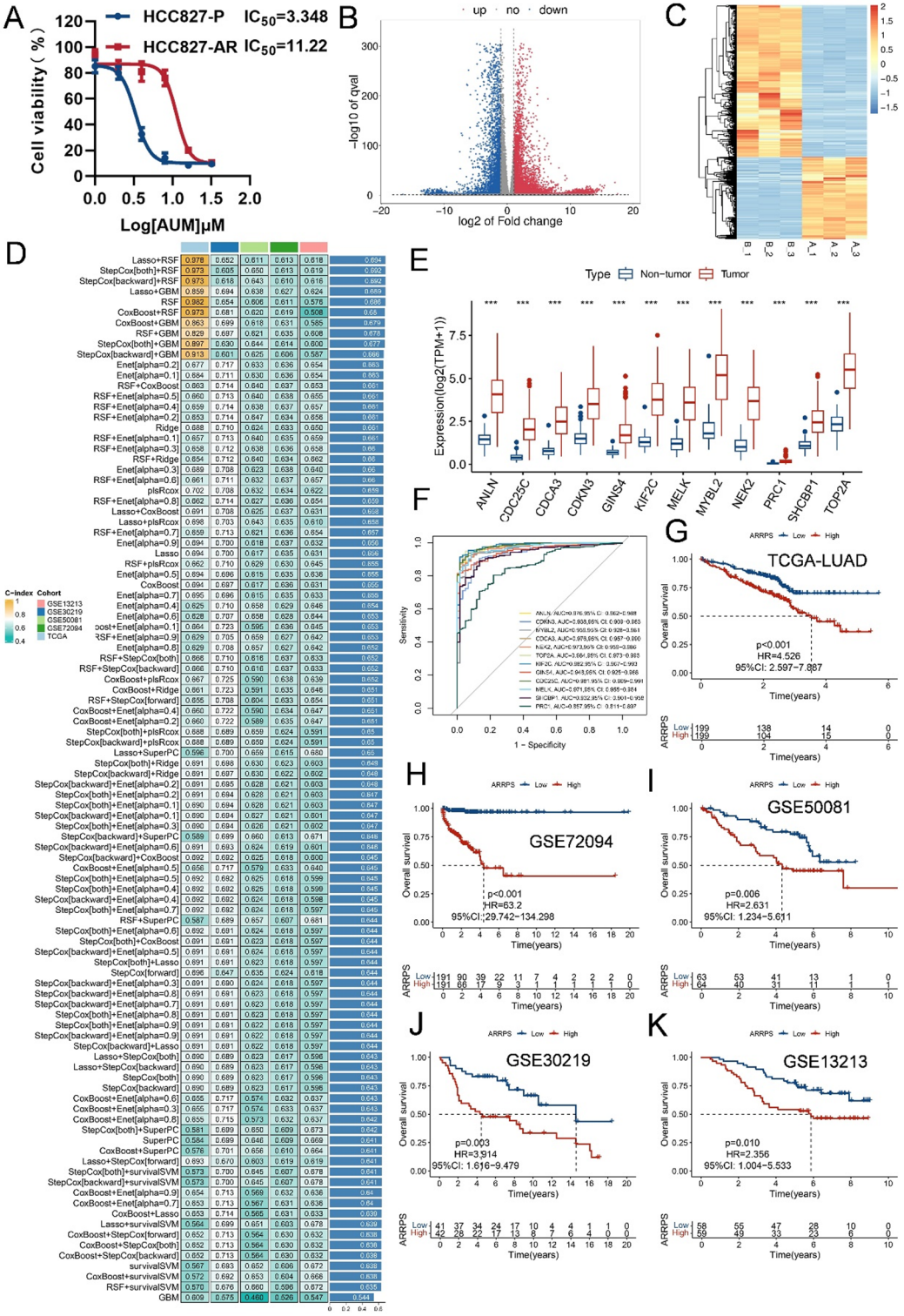

Fig. 2. (**A**) IC50 of HCC827 parental and resistant. B-C. RNA sequencing of HCC827 parental and resistant, differential genes were represented using volcano (**B**) and heat maps (**C**). (**D**) The c-index of 100 combinations of machine learning algorithms in 5 test cohorts. (**E**) Expression of 12 genes in LUAD. (**F**) Diagnostic ROC curves for 12 genes. (**G-K**) Kaplan-Meier analysis of TCGA-LUAD, GSE72094, GSE50081, GSE30219 and GSE13213 datasets.

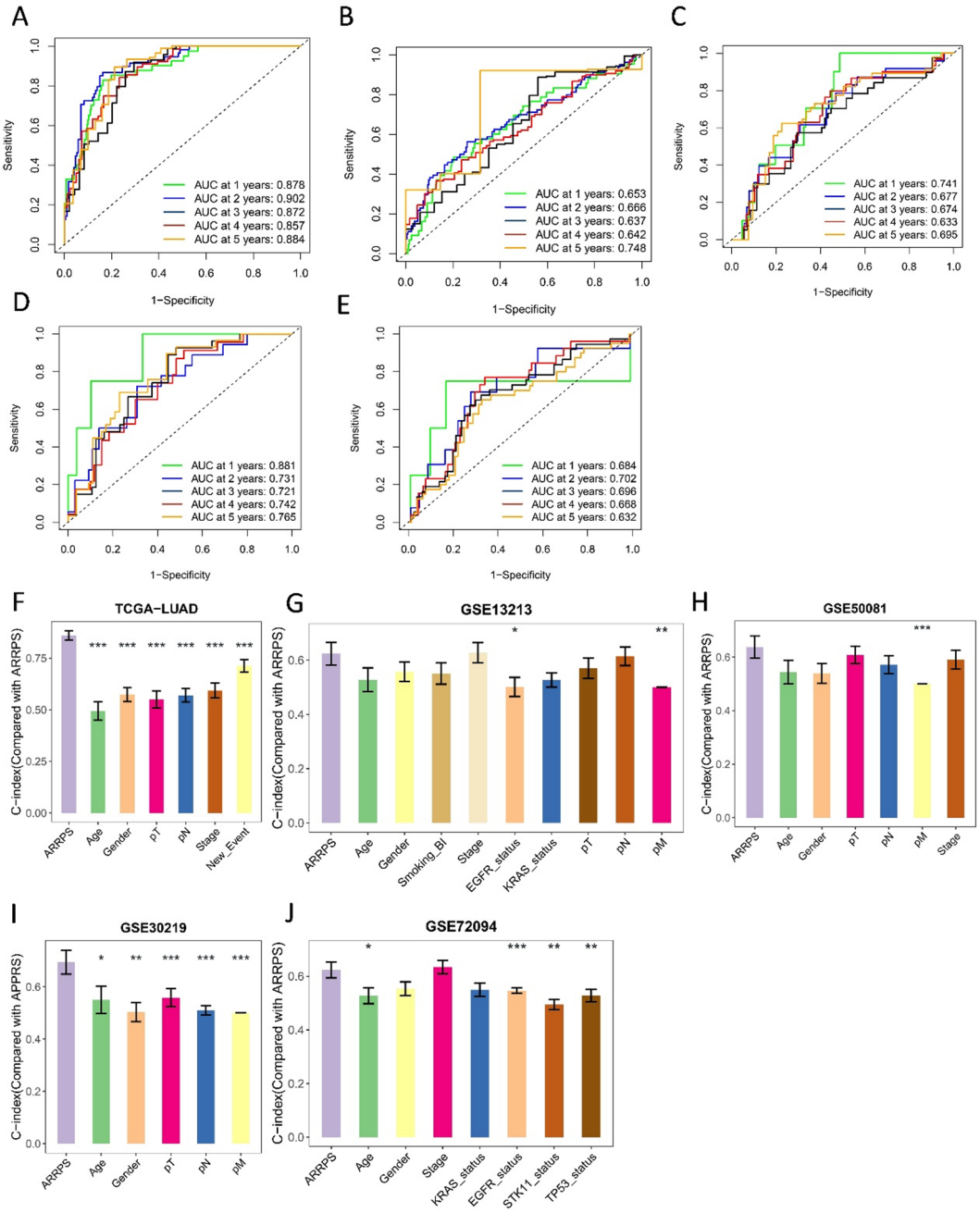

Fig. 3. (**A**-**E**) ROC curve analysis of ARRPS over time in the TCGA-LUAD, GSE72094, GSE50081, GSE30219, and GSE13213 cohorts. (**F**-**J**) C-index when ARRPS is compared with the ability of multiple clinical characteristics to predict OS in LUAD patients.

Differences in pathway enrichment between high and low ARRPS

Signaling pathway enrichment shows the differences involved in the groups with high and low ARRPS feature scores, and the results indicate that the group with high ARRPS feature scores mainly focuses on many signaling pathways that promote tumor growth, such as cell cycle and DNA damage repair and metabolism-related pathways[22–24]such as KEGG_DRUG_METABOLISM_CYTOCHROME_P450, KEGG_ARACHIDONIC_ACID_METABOLISM, KEGG_LINOLEIC_ACID_METABOLISM, KEGG_ETHER_LIPID_METABOLISM, KEGG_ALPHA_LINOLENIC_ACID_METABOLISM, and KEGG_TAURINE_AND_HYPOTAURINE_METABOLISM (Fig. 5B-E).

Analysis of genomic variation between patients in the high and low ARRPS groups

Study of the TCGA-LUAD SNV data found significant differences in SNV between patients in the high and low ARRPS groups. Patients in the high ARRPS group had higher mutation frequencies in NPAP1, TSHZ3, PCDHB8, PTPRT, GRM8, ANK1, KLHL4, TRPC4, COL4A2, LAMA4, and BCHE, whereas patients in the

Predictor	TCGA-LUAD C-index (95% CI)	GSE72094 C-index (95% CI)	GSE50081 C-index (95% CI)	GSE30219 C-index (95% CI)	GSE13213 C-index (95% CI)
ARRPS	**0.73** (0.68–0.78)*	**0.69** (0.63–0.75)*	**0.71** (0.65–0.77)*	**0.67** (0.61–0.73)*	**0.66** (0.60–0.72)*
AJCC Stage (I–IV)	0.68 (0.63–0.73)	0.65 (0.59–0.71)	0.66 (0.60–0.72)	0.62 (0.56–0.68)	0.61 (0.55–0.67)
TP53 Mutation Status	0.62 (0.57–0.67)	0.60 (0.54–0.66)	0.59 (0.53–0.65)	0.58 (0.52–0.64)	0.57 (0.51–0.63)
Smoking Pack-Years	0.59 (0.54–0.64)	0.58 (0.52–0.64)	0.57 (0.51–0.63)	0.56 (0.50–0.62)	0.55 (0.49–0.61)
KRAS Mutation	0.58 (0.53–0.63)	0.57 (0.51–0.63)	\	\	\
EGFR Mutation	0.61 (0.56–0.66)	\	0.60 (0.54–0.66)	\	\
STK11 Mutation	0.57 (0.52–0.62)	\	\	\	\
Age (>65 vs. ≤65)	0.55 (0.50–0.60)	0.54 (0.48–0.60)	0.53 (0.47–0.59)	0.52 (0.46–0.58)	0.51 (0.45–0.57)
Gender	0.53 (0.48–0.58)	0.52 (0.46–0.58)	0.51 (0.45–0.57)	0.50 (0.44–0.56)	0.49 (0.43–0.55)
New Events†	\	0.56 (0.50–0.62)	\	0.54 (0.48–0.60)	\

Table 1. Comparative performance of ARRPS versus clinical features in predicting overall survival (OS) across Cohorts. Bold: Highest C-index per cohort; $P < 0.001$ vs. all other predictors (DeLong's test); †: New tumor events (recurrence/metastasis); \: Data not available in cohort; C-index interpretation: 0.5 = random chance; 1.0 = perfect prediction; Statistical analysis: 95% CI calculated via bootstrap (1,000 replicates); pairwise comparisons via DeLong's test.

low ARRPS group had higher mutation frequencies in CPS1, COL19A1, PLCB4, etc. In terms of CNV, several loci had high CNV frequencies in patients in the high ARRPS group, including 3q26.2(Amp), 12q15(Amp), 17p13.3(Del), 17p11.2(Del), 6q26(Del), 18q22.2(Del), 18q21.1(Del), 22q13.32(Del), 21q21.1(Del), 1p13.1(Del), and 1p36.11(Del) (Fig. S5A). ZBTB4, XAF1, ZMYM2, FOXO1, TNFSF11, ZNF236, BCL2, MAP2K3RBFA, PIK3C3, RBX1, ZNF24, NCF4, SLC5A1 CNV events with more frequencies in patients in the high ARRPS group (Fig. S2A). In addition, we identified multiple gene pairs with SNV co-occurence, including CDH10-MYLK, MRC2-NPAP1, and KIF1A-TSHZ3 (Fig. S3A).

CD-437 and TPCA-1 are drug candidates for patients with high ARRPS

Patients with a high ARRPS have a poorer prognosis and a higher risk of developing resistance to AUM. In order to be able to use the ARRPS as a reliable prognostic risk assessment toolkit, we performed a screening of effective therapeutic agents. The drug screening flowchart is shown in (Fig. 6A). First, differential drug response analysis was performed on the two groups of patients with high and low ARRPS scores to screen compounds of high ARRPS scores patients with low AUC values, and then Spearman correlation analysis was performed to screen compounds with significant negative correlation between AUC values and ARRPS ($r < -0.3$, $Log_2FC < -0.01$), and finally 15 compounds were screened, including 6 CTRP derived compounds (3-Cl-AHPC, cabozantinib, CD-437, mitomycin, nakiterpiosinand StemRegenin 1) and 9 PRISM-derived compounds (broxyquinoline, combretastatin-A-4, deferasirox, frentizole, imiquimod, napabucasin, norepinephrine, tanespimycin, and TPCA-1)(Fig. 6B-E). Finally, by CMap database and searching the previous publications, we found that CD-437 and TPCA-1, with high CMap scores and few studies in LUAD, are drugs with high potential for the treatment of LUAD (Fig. 6F-G). Both CD-437 and TPCA-1 alone significantly inhibited the viability of drug-resistant cell lines but were less sensitive to parental (Fig. S6 A-B). We then sought to combine both with AUM to explore the sensitivity of drug-resistant cell lines to AUM. We used the CCK8 assay to detect the antiproliferative effects of CD-437 and TPCA-1 in combination with AUM on drug-resistant cell lines and expressed them as a dose-response matrix and a synergy scoring matrix (Fig. 7A-C). Synergy scores were calculated by SynergyFinder using the HSA model[25,26]. It was found that this combination exerted a stronger antiproliferative effect. The same therapeutic effect was subsequently detected using clone formation (Fig. 7D), not also. In addition to that, this combination effect significantly inhibited the growth of xenograft tumors in nude mice (Fig. 7E-J, Fig. S6C). In conclusion, these results of ours surface that the two compounds screened by high and low ARRPS can effectively inhibit the growth of NSCLC.

Discussion

Resistance to EGFR-TKI has long been a major impediment to targeted therapy for EGFR-mutant NSCLC[27–29]. The conventional solution is to increase the dose or add additional other chemotherapeutic agents. However, this is not a good solution. Immunotherapy has a more expensive cost and serious adverse effects. Therefore, exploring a new biomarker to effectively treat EGFR mutant NSCLC is also necessary. To achieve this goal, we built multiple prognostic models based on AUM resistance-associated differential genes using a combination of 10 machine learning algorithms. Subsequently, the 12-gene ARRPS, derived from the combination of lasso + GBM, emerged as superior.

As a linear regression model, Lasso regression selects features during model training and can be used for feature selection and dimensionality reduction. It also improves the generalization ability of the model and avoids overfitting. However, in this study, its direct application of this model produces not much accuracy, achieving an average c-index of 0.655. GBM has the advantages of good interpretability and wide applicability, but it lacks the inherent ability to downscale and carries the risk of overfitting. To address these challenges, we combine the advantages of CoxBoost and GBM. It turns out that this combined approach proves to be the most effective, outperforming other models in our study.

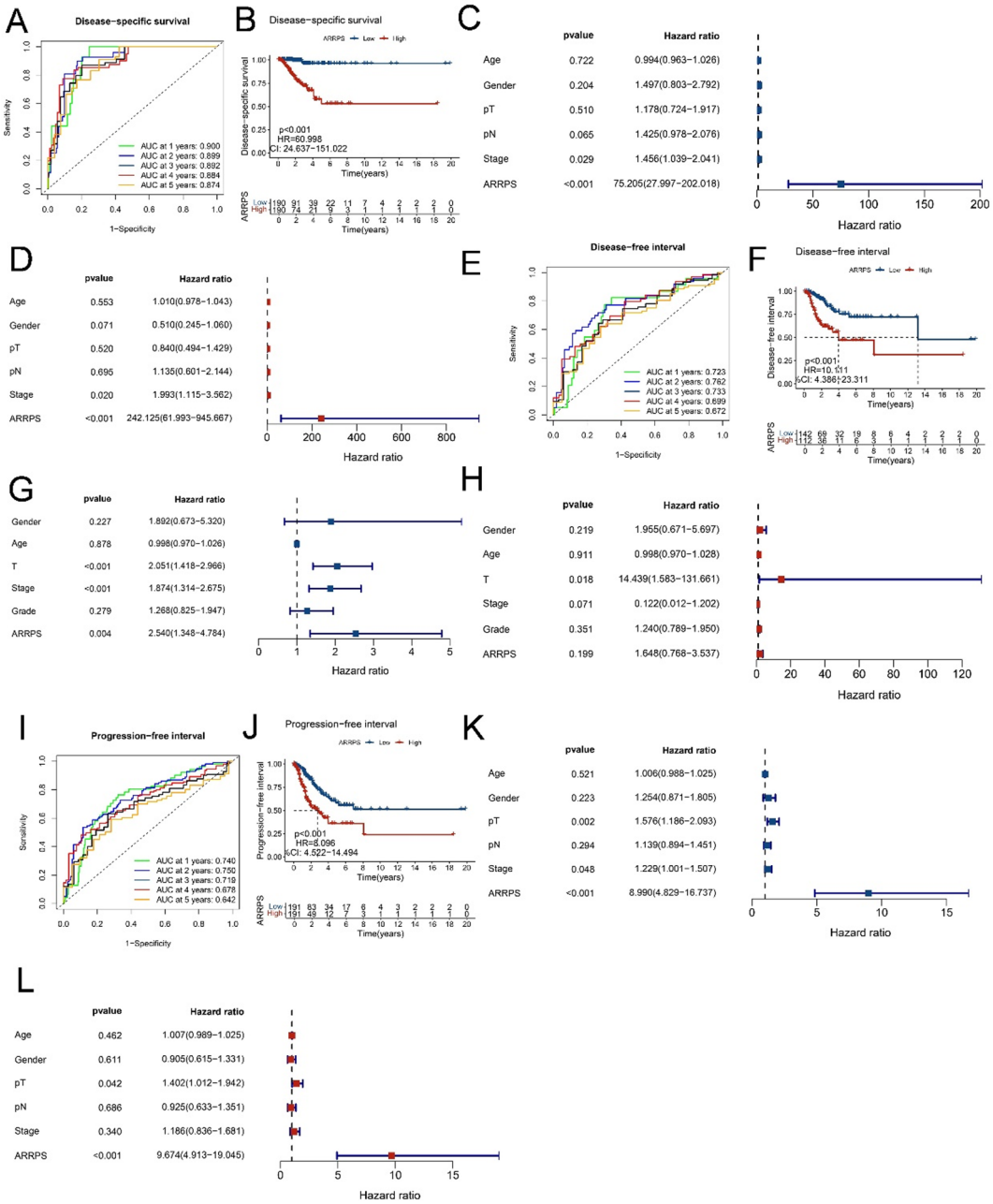

Fig. 4. (**A**-**L**) The relationship between ARRPS and DSS was analyzed through Roc curves, Kaplan-Meier, univariate COX regression, and multivariate COX regression. (**A**-**D**) The relationship between ARRPS and DFI was analyzed through Roc curves, Kaplan-Meier, univariate COX regression, and multivariate COX regression (**E**-**H**). The relationship between ARRPS and PFI was analyzed through Roc curves, Kaplan-Meier, univariate COX regression, and multivariate COX regression (**I**-**L**).

In our study, ARRPS was found to be a reliable prognostic risk assessment tool. This also gives us a greater interest in the screening of therapeutic agents for patients with high ARRPS. It revealed that CD-437 and TPCA-1 had good inhibitory ability against drug-resistant cell lines, but not parental cells (Fig. S6 A-B). Then we combined them with AUM, which also showed good synergistic anti-tumor ability, both in vivo and in vitro. In clinical practice, we can obtain biopsy samples from diagnosed NSCLC patients and perform RNA-seq analysis to elucidate their transcriptome profiles. Subsequently, we can reconcile these profiles with the meta-LUAD cohort and calculate the ARRPS. Based on the ARRPS threshold, NSCLC patients can be categorized into a low ARRPS group and a high ARRPS group. Conventional therapy is sufficient for patients in the low ARRPS group, while those in the high ARRPS group may benefit from the combination of AUM with CD-437 and TPCA-1.

The development of ARRPS addresses a critical unmet need in managing acquired resistance to third-generation EGFR TKIs—a phenomenon that ultimately afflicts most advanced NSCLC patients. While our signature originated from an AUM-resistant HCC827 model, its validation across five independent LUAD

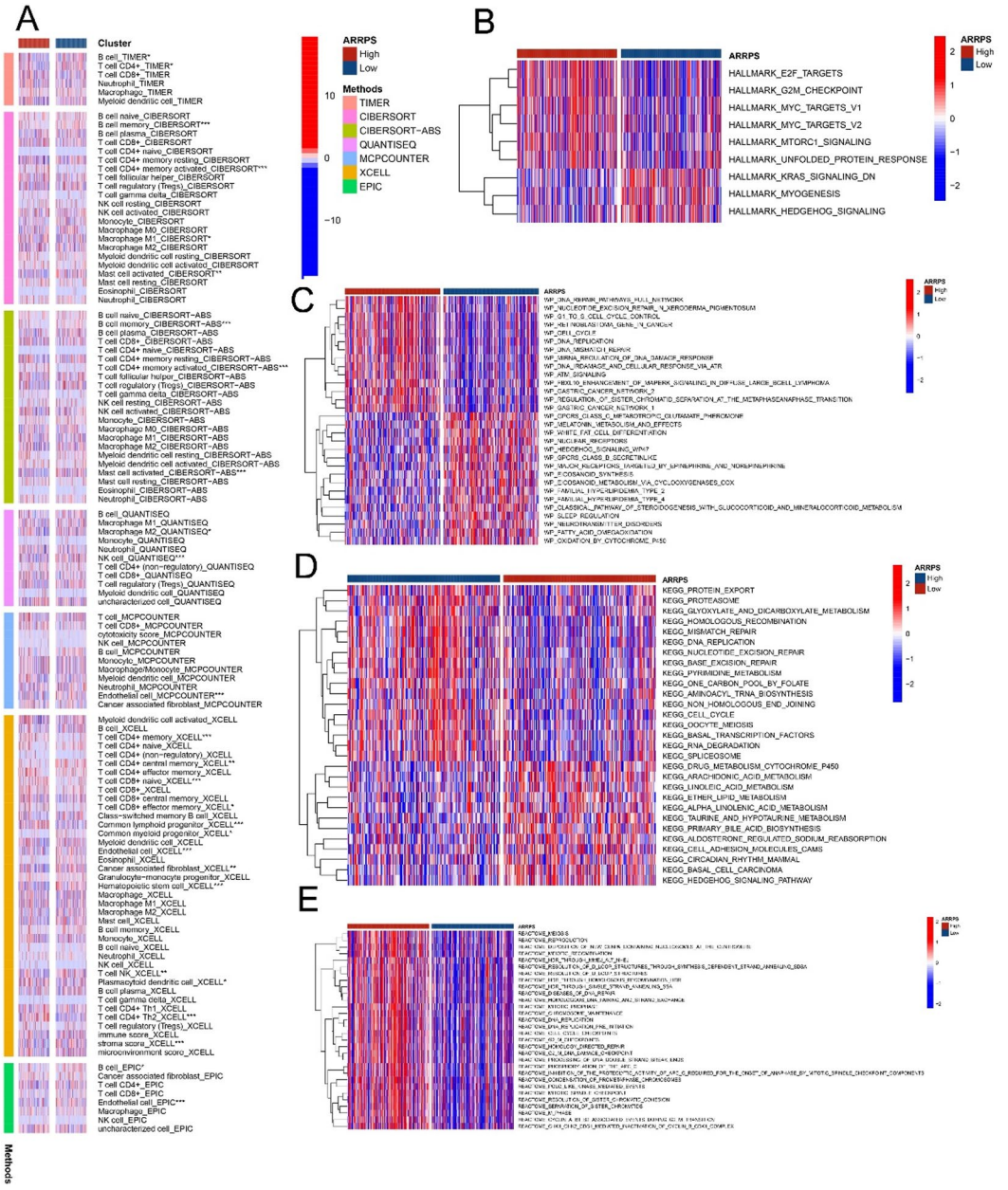

Fig. 5. (**A**) Differences in immune background between high and low ARRPS scores. (**B-E**) Differences in pathway enrichment between high and low ARRPS scores[22–24].

cohorts ($n = 1{,}412$) demonstrates robust generalizability to clinical resistance heterogeneity. Crucially, ARRPS outperformed 40 published signatures and traditional prognosticators like AJCC staging, suggesting it captures fundamental biological drivers of TKI failure beyond cell line artifacts. Notably, the 12-gene panel (e.g., FOXM1, CDC20, CENPF) encodes master regulators of mitotic fidelity and DNA damage response—pathways recurrently altered in osimertinib-resistant tumors via AURKA-driven evolutionand phenotypic transformation[6,28]. This mechanistic conservation implies ARRPS's applicability to global TKI resistance landscapes, a hypothesis currently under validation in osimertinib-resistant PDX models. Furthermore, the therapeutic relevance of CD-437 and TPCA-1 extends beyond AUM-specific contexts. Critically, both agents exhibited 4–6-fold selectivity for resistant versus parental cells, confirming their targeting of acquired dependencies rather than generic cytotoxicity. Their synergy with AUM (HSA scores > 10) further supports a complementary mechanism: TKIs suppress primary EGFR signaling, while CD-437/TPCA-1 override resistance-specific survival adaptations.

However, this study also has potential limitations. Key barriers include the cost and accessibility of RNA-seq for ARRPS scoring, limited availability of clinical-grade CD-437/TPCA-1, and regulatory barriers to combination therapies. To mitigate these issues, we propose developing a targeted 12-gene NanoString assay (validation underway) that could reduce costs and is suitable for CLIA-accredited laboratories. Regarding drug

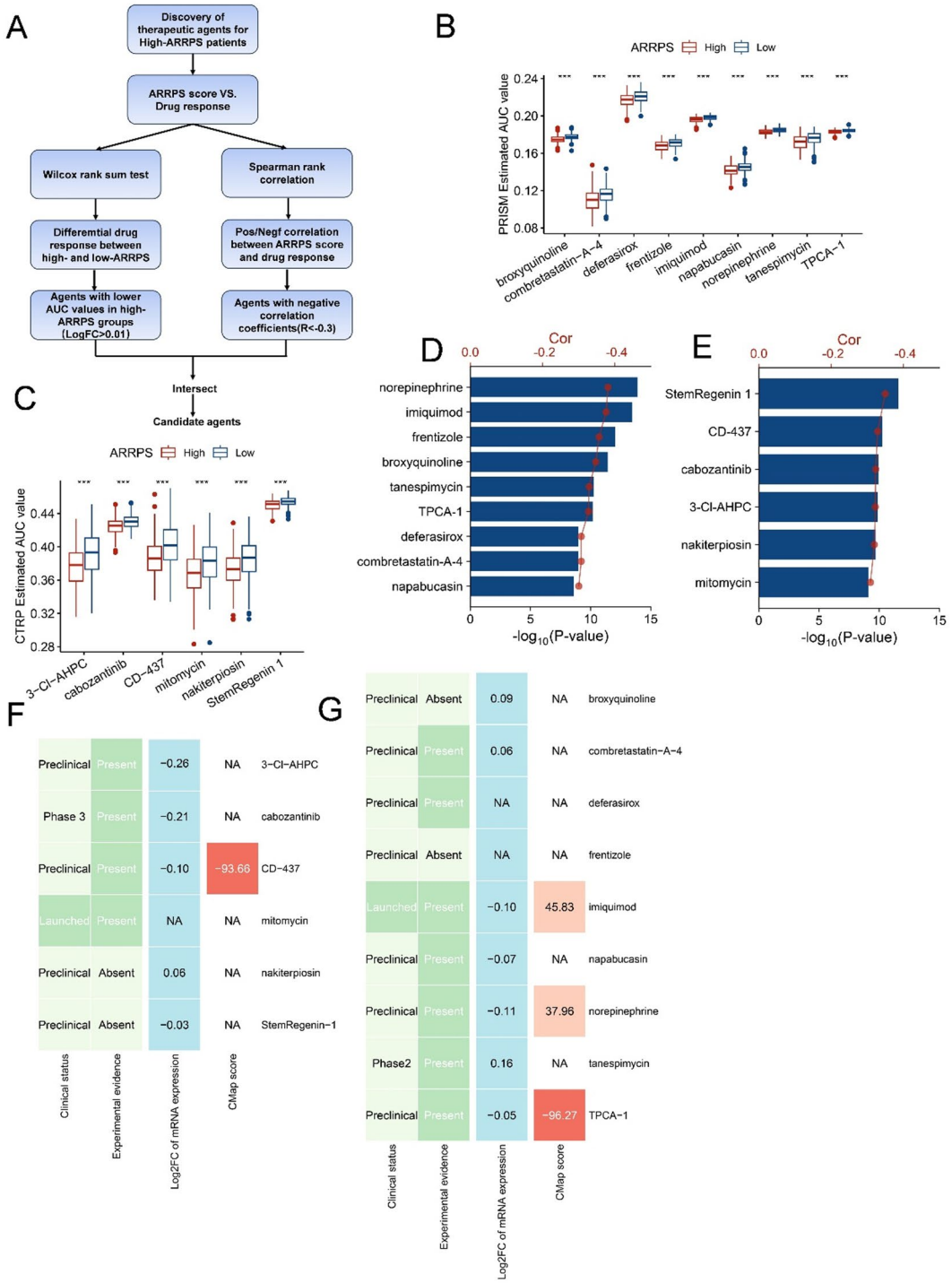

Fig. 6. (**A**) Schematic representation of the strategy to marshal potential therapeutic agents with higher drug sensitivity in the high ARRPS group. (**B**-**C**) Differential drug response analysis of Prism-derived compounds (**B**) and CTRP-derived compounds (**C**) showed that. (**D**-**E**) Results of Spearman correlation analysis of the relative inhibitory effects of CTRP (**D**) and PRISM-derived compounds (**E**). (**F**-**G**) Screening of thera-peutic agents by CMap.

accessibility, we are exploring FDA-approved retinoids (e.g., adapalene) as CD-437 analogs and working with manufacturers to produce GMP-grade TPCA-1.

In summary, this study introduced ARRPS as a potential tool for identifying high-risk subgroups of patients with LUAD and identified therapeutic candidates CD-437 and TPCA-1 targeting high ARRPS (Fig. S7), and we summarise the key findings in Table 2. These results provide a promising therapeutic avenue and represent a potential complementary treatment for high-risk LUAD patients.

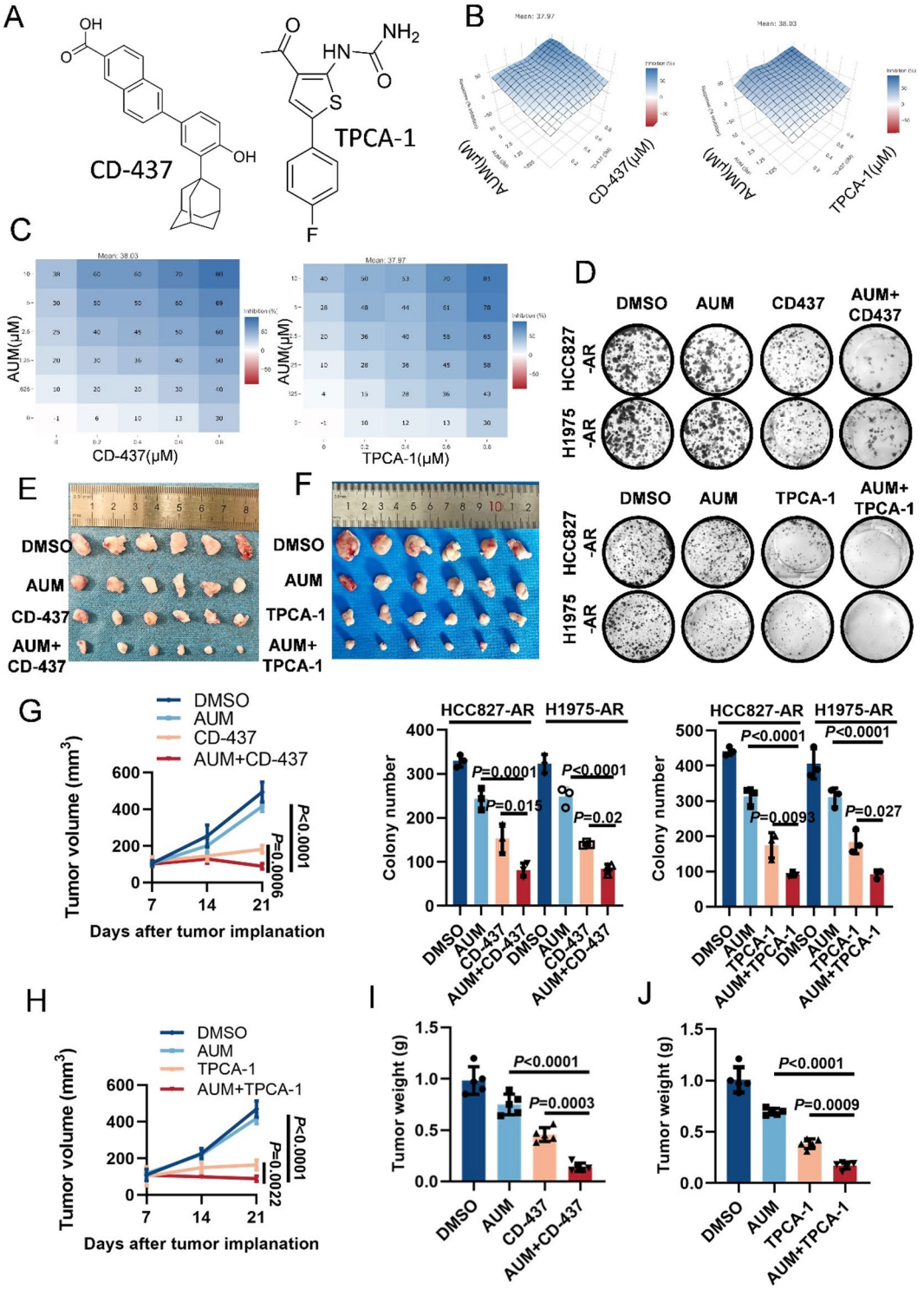

Fig. 7. (**A**) Schematic chemical structures of CD-437 and TPCA-1. (**B-C**) The effects of CD-437 and TPCA-1 in combination with AUM on cell viability were examined using CCK8 and are shown as a dose-response matrix (**B**) and synergism score matrix (**C**). (**D**) Effects on colony formation under combination therapy. (**E-F**) Antitumor effects of CD-437 (**E**) and TPCA-1 (**F**) in combination with AUM including. (**G-H**) Tumor volume changes. (**I-J**) Tumor tissue weight.

Category	Core findings	Validation methods	Clinical significance
Prognostic model (ARRPS)	12-gene signature (ANLN, ASPM, BUB1B, CDC20, CDC25C, CDC6, CDCA2, CDCA3, CDKN3, CENPE, CENPF, FOXM1)	10 ML algorithms × 5 independent cohorts	Superior OS/PFI prediction vs. AJCC staging and TP53 mutations
Prognostic performance	High ARRPS group: 3.2× higher mortality risk (avg. HR = 3.2, $P < 0.001$)	KM survival analysis + C-index (0.72–0.81)	Identifies high-risk patients for intensified therapy
Resistance mechanisms	(1) Upregulated cell cycle/DNA repair pathways (FOXM1, TOP2A) (2) Metabolic reprogramming (fatty acid/ether lipid metabolism)	GSVA pathway enrichment + SNV/CNV analysis.	Reveals biological basis of resistance; suggests therapeutic targets
Therapeutic strategy	(1) CD-437 + AUM: synergistically inhibits resistant cells (Synergy Score > 20) (2) TPCA-1 + AUM: suppresses xenograft growth ($P < 0.01$)	In vitro drug sensitivity + in vivo xenograft experiments.	Provides personalized combo therapy for high ARRPS patients.

Table 2. Key findings summary.

Data availability

The datasets generated and/or analyzed during the current study are available in the NCBI BioProject repository under accession number PRJNA1248195 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1248195). The data sets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 26 March 2025; Accepted: 12 September 2025
Published online: 28 October 2025

References

1. Passaro, A., Jänne, P. A., Mok, T. & Peters, S. Overcoming therapy resistance in EGFR-mutant lung cancer. *Nat. Cancer*. **2**, 377–391. https://doi.org/10.1038/s43018-021-00195-8 (2021).
2. Zhang, T. et al. Correction: Discovery of a novel third-generation EGFR inhibitor and identification of a potential combination strategy to overcome resistance. *Mol. Cancer*. **22**, 139. https://doi.org/10.1186/s12943-023-01842-7 (2023).
3. Pan, Z. et al. Cholesterol promotes EGFR-TKIs resistance in NSCLC by inducing EGFR/Src/Erk/SP1 signaling-mediated ERRα re-expression. *Mol. Cancer*. **21**, 77. https://doi.org/10.1186/s12943-022-01547-3 (2022).
4. Shih, P. C. The role of the STAT3 signaling transduction pathways in radioresistance. *Pharmacol. Ther.* **234**, 108118. https://doi.org/10.1016/j.pharmthera.2022.108118 (2022).
5. Liu, Z. et al. Machine learning-based integration develops an immune-derived LncRNA signature for improving outcomes in colorectal cancer. *Nat. Commun.* **13**, 816. https://doi.org/10.1038/s41467-022-28421-6 (2022).
6. Shah, K. N. et al. Aurora kinase A drives the evolution of resistance to third-generation EGFR inhibitors in lung cancer. *Nat. Med.* **25**, 111–118. https://doi.org/10.1038/s41591-018-0264-7 (2019).
7. Liu, Z. et al. Integrative analysis from multi-center studies identities a consensus machine learning-derived LncRNA signature for stage II/III colorectal cancer. *EBioMedicine* **75**, 103750. https://doi.org/10.1016/j.ebiom.2021.103750 (2022).
8. Yang, Z. et al. YKT6, as a potential predictor of prognosis and immunotherapy response for oral squamous cell carcinoma, is related to cell invasion, metastasis, and CD8 + T cell infiltration. *Oncoimmunology* **10**, 1938890. https://doi.org/10.1080/2162402x.2021.1938890 (2021).
9. Wang, L. et al. Multi-omics landscape and clinical significance of a SMAD4-driven immune signature: implications for risk stratification and frontline therapies in pancreatic cancer. *Comput. Struct. Biotechnol. J.* **20**, 1154–1167. https://doi.org/10.1016/j.csbj.2022.02.031 (2022).
10. Tan, Z. et al. The value of a metabolic reprogramming-related gene signature for pancreatic adenocarcinoma prognosis prediction. *Aging (Albany NY)*. **12**, 24228–24241. https://doi.org/10.18632/aging.104134 (2020).
11. Yuan, Q. et al. Development and validation of a novel N6-methyladenosine (m6A)-related multi- long non-coding RNA (lncRNA) prognostic signature in pancreatic adenocarcinoma. *Bioengineered* **12**, 2432–2448. https://doi.org/10.1080/21655979.2021.1933868 (2021).
12. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A*. **102**, 15545–15550. https://doi.org/10.1073/pnas.0506580102 (2005).
13. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–w514. https://doi.org/10.1093/nar/gkaa407 (2020).
14. Chrétien, S. et al. PD-1/PD-L1 inhibition: what the future holds for breast cancer immunotherapy. *Cancers (Basel)*. 11. https://doi.org/10.3390/cancers11050628 (2019).
15. Wang, J. et al. Fibrinogen-like protein 1 is a major immune inhibitory ligand of LAG-3. *Cell* **176**, 334–347e312. https://doi.org/10.1016/j.cell.2018.11.010 (2019).
16. Janakiram, M., Chinai, J. M., Zhao, A., Sparano, J. A. & Zang, X. HHLA2 and TMIGD2: new immunotherapeutic targets of the B7 and CD28 families. *Oncoimmunology* **4**, e1026534. https://doi.org/10.1080/2162402x.2015.1026534 (2015).
17. Bindea, G. et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795. https://doi.org/10.1016/j.immuni.2013.10.003 (2013).
18. Yang, C. et al. A survey of optimal strategy for signature-based drug repositioning and an application to liver cancer. *Elife* **11**. https://doi.org/10.7554/eLife.71880 (2022).
19. Yang, K. et al. Complement map database. *Bioinformatics* **29**, 1832–1833. https://doi.org/10.1093/bioinformatics/btt269 (2013).
20. Maeser, D., Gruener, R. F. & Huang, R. S. OncoPredict: An R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* **22**. https://doi.org/10.1093/bib/bbab260 (2021).
21. Geeleher, P., Cox, N. & Huang, R. S. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One*. **9**, e107468. https://doi.org/10.1371/journal.pone.0107468 (2014).
22. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. https://doi.org/10.1002/pro.3715 (2019).
23. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–d677. https://doi.org/10.1093/nar/gkae909 (2025).
24. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
25. Ianevski, A., Giri, A. K. & Aittokallio, T. SynergyFinder 3.0: an interactive analysis and consensus interpretation of multi-drug synergies across multiple samples. *Nucleic Acids Res.* **50**, W739–w743. https://doi.org/10.1093/nar/gkac382 (2022).

26. Zheng, S. et al. SynergyFinder plus: toward better interpretation and annotation of drug combination screening datasets. *Genomics Proteom. Bioinf.* **20**, 587–596. https://doi.org/10.1016/j.gpb.2022.01.004 (2022).
27. Fu, Y. et al. Abnormally activated opn/integrin αVβ3/FAK signalling is responsible for EGFR-TKI resistance in EGFR mutant non-small-cell lung cancer. *J. Hematol. Oncol.* **13**, 169. https://doi.org/10.1186/s13045-020-01009-7 (2020).
28. Nilsson, M. B. et al. CD70 is a therapeutic target upregulated in EMT-associated EGFR tyrosine kinase inhibitor resistance. *Cancer Cell.* **41**, 340–355e346. https://doi.org/10.1016/j.ccell.2023.01.007 (2023).
29. Peng, S. et al. EGFR-TKI resistance promotes immune escape in lung cancer via increased PD-L1 expression. *Mol. Cancer*. **18**, 165. https://doi.org/10.1186/s12943-019-1073-4 (2019).

Author contributions
Xiao Wu performed most of the experiments, Yongxia Chen and Guolei Song analyzed data, Yang Lu, Yongping Li and Annan Zhu contributed to experiments. Tingting Wang, Zishu Wang and Fang Su designed experiments, supervised the study, and wrote the manuscript. All authors consent to publication of this article.

Funding
This study was supported by Natural Science key project of Bengbu Medical College (No. 2021byzd064).

Declarations

Ethics approval and consent to participate
The animal experiments were conducted in accordance with ARRIVE 2.0 guidelines (Animal Research: Reporting of In Vivo Experiments, https://arriveguidelines.org). All animal experimental procedures were approved by the Experimental Animal Teaching and Research Committee of Bengbu Medical University (2023 – 522). All methods were performed in accordance with the relevant guidelines and regulations.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-20159-7.

Correspondence and requests for materials should be addressed to G.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.