

# Development and validation of a novel epigenetic-related prognostic signature and candidate drugs for patients with lung adenocarcinoma

Zhihao Wang<sup>1</sup>, Kidane Siele Embaye<sup>1</sup>, Qing Yang<sup>2</sup>, Lingzhi Qin<sup>1</sup>, Chao Zhang<sup>1</sup>, Liwei Liu<sup>1</sup>, Xiaoqian Zhan<sup>1</sup>, Fengdi Zhang<sup>3</sup>, Xi Wang<sup>1</sup>, Shenghui Qin<sup>1</sup>

<sup>1</sup>Institute of Pathology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

<sup>2</sup>Department of Pharmacy, Hiser Medical Center of Qingdao, Qingdao 266033, China

<sup>3</sup>Department of Pathology, Wuhan Third Hospital, Tongren Hospital of Wuhan University, Wuhan 430030, China

**Correspondence to:** Shenghui Qin; **email:** [2015tj0147@hust.edu.cn](mailto:2015tj0147@hust.edu.cn)

**Keywords:** lung adenocarcinoma, epigenetic-related genes, drugs, prognostic, TCGA

**Received:** November 9, 2020

**Accepted:** May 11, 2021

**Published:** July 20, 2021

**Copyright:** © 2021 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Background:** Epigenetic dysregulation has been increasingly proposed as a hallmark of cancer. Here, the aim of this study is to establish an epigenetic-related signature for predicting the prognosis of lung adenocarcinoma (LUAD) patients.

**Results:** Five epigenetic-related genes (ERGs) (ARRB1, PARP1, PKM, TFDP1, and YWHAZ) were identified as prognostic hub genes and used to establish a prognostic signature. According our risk score system, LUAD patients were stratified into high and low risk groups, and patients in the high risk group had a worse prognosis. ROC analysis indicated that the signature was precise in predicting the prognosis. A new nomogram was constructed based on the five hub genes, which can predict the OS of every LUAD patients. The calibration curves showed that the nomogram had better accuracy in prediction. Finally, candidate drugs that aimed at hub ERGs were identified, which included 47 compounds.

**Conclusions:** Our epigenetic-related signature nomogram can effectively and reliably predict OS of LUAD patients, also we provide precise targeted chemotherapeutic drugs.

**Methods:** The genomic data and clinical data of LUAD cohort were downloaded from the TCGA database and ERGs were obtained from the EpiFactors database. GSE31210 and GSE50081 microarray datasets were included as independent external datasets. Univariate Cox, LASSO regression, and multivariate Cox analyses were applied to construct the epigenetic-related signature.

## INTRODUCTION

Lung cancer, which including two main types small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), is the most common cause of cancer deaths in the world [1]. NSCLC accounts for ~85% of lung cancer diagnoses and imposes a heavy burden on global health systems [2, 3]. Lung adenocarcinoma (LUAD) is the most common pathological subtype of NSCLC. Despite great advances in the treatment of LUAD, the clinical outcome is not satisfactory [4]. Recently, a number of

molecularly targeted therapies have been developed that caused significant improvement in the treatment of LUAD, especially for patients with EGFR mutation [5] and ALK mutation [6, 7]. However, because of high gene mutation heterogeneity and complexity in molecular patterns of LUAD, large amount of LUAD patients without EGFR and ALK mutations lose the opportunity to use effective therapeutic drugs. Therefore, it is essential to identify reliable diagnostic and prognostic biomarkers for LUAD patients and thereby provide precise targeted therapies.

Epigenetic regulation is broadly defined as repression or activation of gene expression via DNA methylation and histone modification, without introducing changes in the DNA sequence per se [8, 9]. Epigenetic dysregulation is considered as an essential hallmark in the initiation and progression of different types of cancers [10, 11]. Recently, epigenetic dysregulation has been reported to promote carcinogenesis of pulmonary epithelial cells and the progression of LUAD [12]. Epigenetic genes, such as DNMT, HDAC, and PARP1 have emerged as attractive targets for the development of anticancer drugs for NSCLC [13–15]. Hence, epigenetic alterations are considered to be an important characteristic in NSCLC development and progression. Although, numerous studies have reported prognostic signature for LUAD [16–19], only a few researchers have explored the role of epigenetic prognostic markers in LUAD and found out potential drug candidates for targeted therapy.

In the present study, we firstly developed an epigenetic-related prognostic signature based on TCGA dataset of 497 LUAD patients, and then validated it in both GSE31210 and GSE50081 datasets. More importantly, a new nomogram was created for predicting the overall survival (OS) of LUAD patients based on five epigenetic-related genes (ERGs). The accurate prediction function of the nomogram was evaluated. Moreover, we further screened out candidate targeted chemotherapy drugs to the five ERGs. In conclusion, our study may provide new ideas for epigenetic-related prognostic biomarkers, thereby, highlighting the need to identify high risk patients and eventually delivering more effective targeted chemotherapeutic drugs for LUAD patients.

## RESULTS

### Identifying differentially expressed ERGs

The flow diagram of our study is illustrated in Figure 1A. Firstly, we downloaded mRNA data of 497 LUAD samples and 54 corresponding normal lung samples and corresponding clinical information from the TCGA database. Meanwhile, a total of 720 ERGs were obtained from the EpiFactors database. Then, we matched 720 ERGs with LUAD-related mRNA data, 91 differentially expressed ERGs were identified ( $|\log FC| > 1.0$ , adjusted  $P$  value  $< 0.05$ ), including 12 downregulated and 79 upregulated ERGs (Figure 1).

### GO and KEGG enrichment analyses

GO analysis results indicated that 91 ERGs were mainly enriched in 422 Go terms, such as covalent chromatin modification, histone modification, chromatin remodeling, peptidyl-lysine modification, regulation of chromosome organization and so forth (Figure 2A, 2B).

KEGG analysis results demonstrated that 91 ERGs were mainly enriched in 4 pathways including homologous recombination, cell cycle, lysine degradation and fanconi anemia (Figure 2C, 2D).

### Identification of survival-related differentially expressed ERGs

By using univariate Cox regression analysis, we identified 16 ERGs that were significantly associated with prognosis in patients with LUAD (Table 1) ( $P < 0.05$ ; Figure 3A). Of the 16 genes, 14 were identified as risk factors and two were identified as protective factors. UHRF1, HJURP, BUB1, PBK, AURKA, ACTL6A, ASF1B, PRKDC, CDK1, PARP1, UBE2T, YWHAZ, PKM, and TFDP1 were identified as risk factors ( $P < 0.05$ ; HRs, 1.0012-1.0740). Whereas ARRB1 and CBX7 were considered as protective factors ( $P < 0.05$ ; HRs, 0.9512 and 0.8885, respectively). The expression of 16 ERGs in LUAD and normal lung samples were presented in Figure 3B, and the correlation between these ERGs could be seen in Figure 3C. Due to the important prognostic value of the candidate genes, their genetic alterations were analyzed. As shown in Figure 3D, missense mutation is the most common type of mutation, and there are 10 genes with mutation rate  $\geq 3\%$ , among which PRKDC mutation is the most frequent (12%).

### Development of epigenetic-related signature

To eliminate highly correlated ERGs and prevent overfitting of the signature, the Lasso regression analysis was performed (Figure 4A, 4B). Finally, five ERGs were confirmed by multivariate Cox regression analysis and applied to develop an epigenetic-related prognostic signature (Table 1). The signature was established to assess the prognosis of each LUAD patient as follows:

$$\text{Risk score} = (-0.03706 \times \text{expression level of ARRB1}) + (0.015474 \times \text{expression level of PARP1}) + (0.004509 \times \text{expression level of YWHAZ}) + (0.001132 \times \text{expression level of TFDP1}) + (0.004303 \times \text{expression level of PKM}).$$

According to above formula, the risk scores of 426 LUAD patients were calculated, and patients were clustered into high and low risk groups by median risk score. The risk score (Figure 5A) and survival status (Figure 5B) of 426 LUAD patients are presented. As shown in Figure 5C, patients in the high risk group had significantly poorer OS compared to those in the low risk group. The prognostic signature exhibited good predictive performance and accuracy for predicting the 1-, 3- and 5-year OS (Figure 5D).



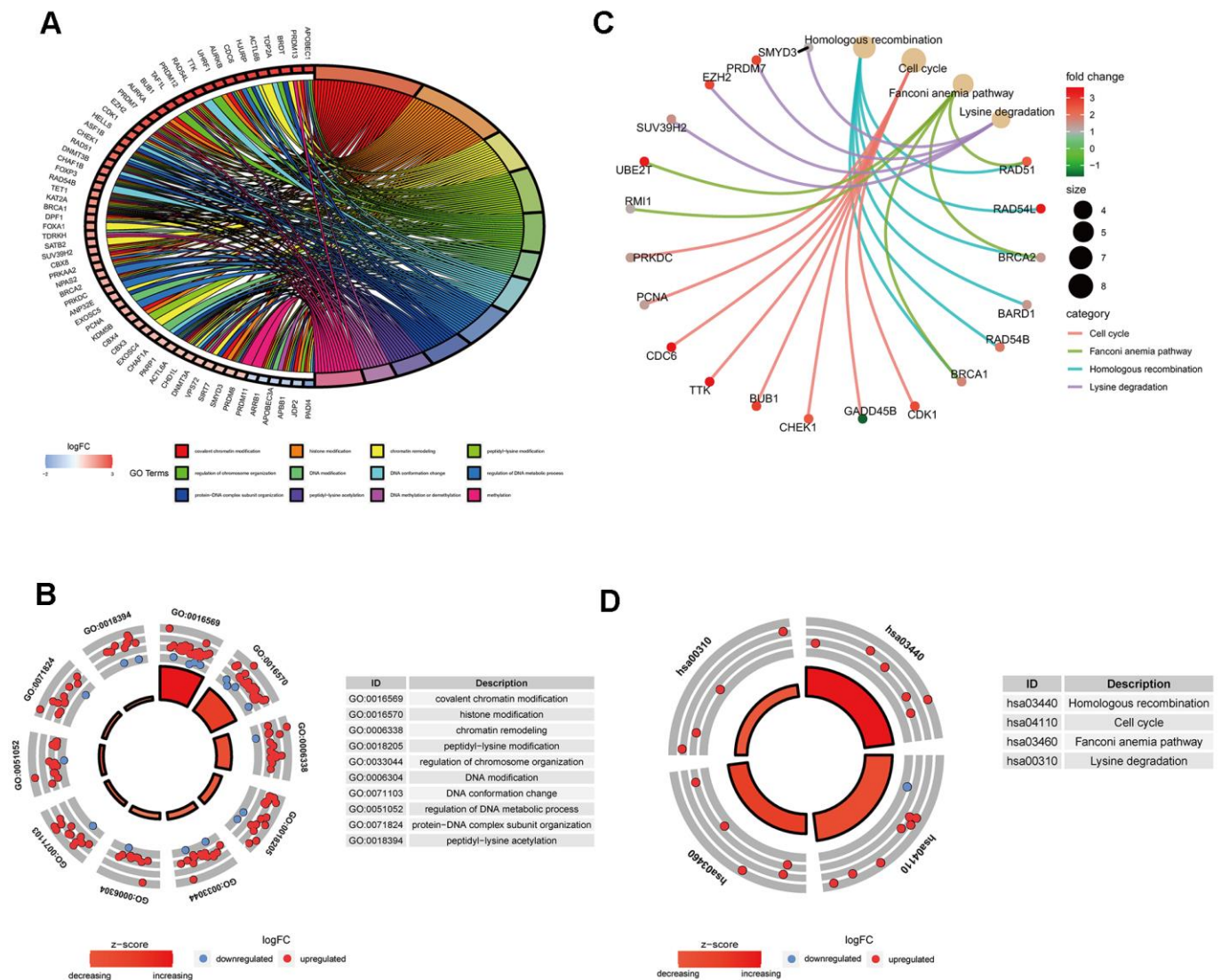


## Relationships between risk score and clinicopathological factors

Furthermore, the expression of the five crucial genes and clinicopathological factors in high and low risk groups are presented in Figure 6A. The results indicated that with reduction of the risk score, ARRB1 was gradually increased, while YWHAZ, PKM, PARP1, and TFDP1 were gradually decreased. Significant differences were found for the pathologic T stage between the high and low risk groups. Moreover, a higher risk score was associated with higher pathological stage, N-stage and some histological growth patterns, including adenocarcinoma, adenocarcinoma with mixed subtypes,

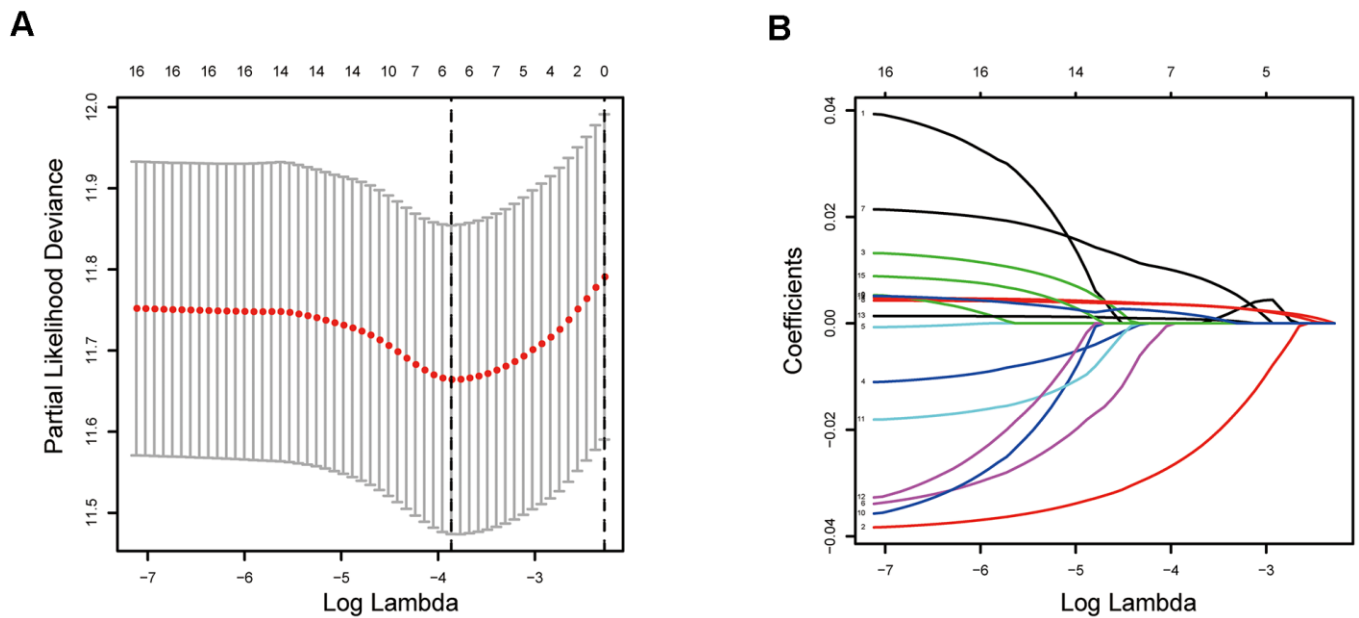
bronchioloalveolar carcinoma, non-mucinous, and papillary adenocarcinoma (Figure 6B).

Next, Univariate Cox analysis data demonstrated that pathologic stage, T stage, N stage and the risk score were all associated with OS (Figure 6C). Multivariate analysis data demonstrated that stage, M and the risk score ( $P < 0.001$ ) could be used as an independent prognostic factors (Figure 6D). Moreover, YWHAZ is significantly correlated with pathological stage, PKM is significantly correlated with pathological stage and N stage ( $P < 0.05$ , Table 2). Then the relationship between the risk score and clinicopathological factors was assessed. The results indicated that patients with high risk scores were

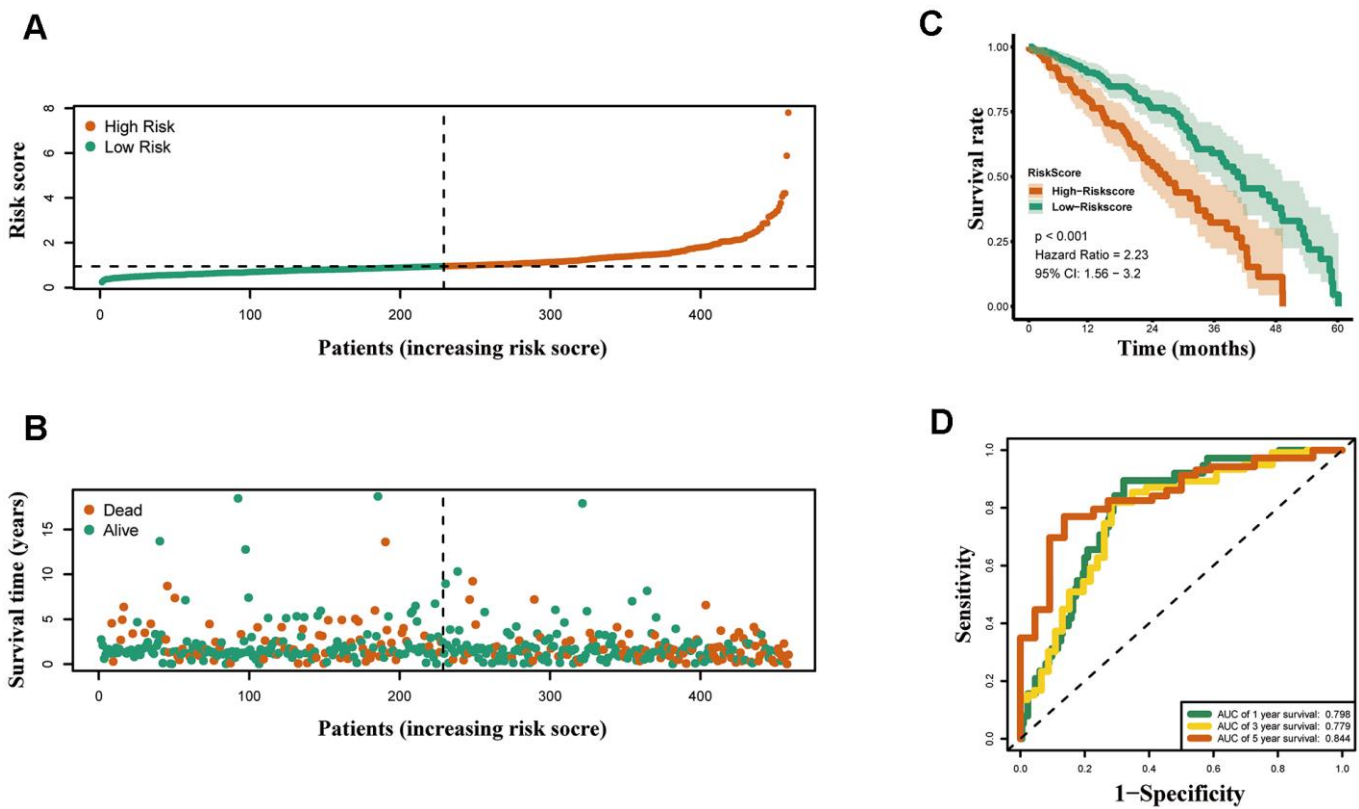


**Figure 2. Gene functional enrichment analysis of differentially expressed ERGs.** (A) The top 12 significant terms of GO function enrichment. (B) The GO circle shows the scatter map of the logFC of the specified gene. (C) The correlation between intersection genes and the significant terms of KEGG. (D) The KEGG circle shows the scatter map of the logFC of the specified gene. The higher the Z-score value indicated, the higher expression of the enriched pathway.





**Figure 4. Establishment of epigenetic-related prognostic signature. (A)** Screening of optimal parameter (lambda) at which the vertical lines were drawn. **(B)** Lasso coefficient profiles of these ERGs with non-zero coefficients determined by the optimal lambda.



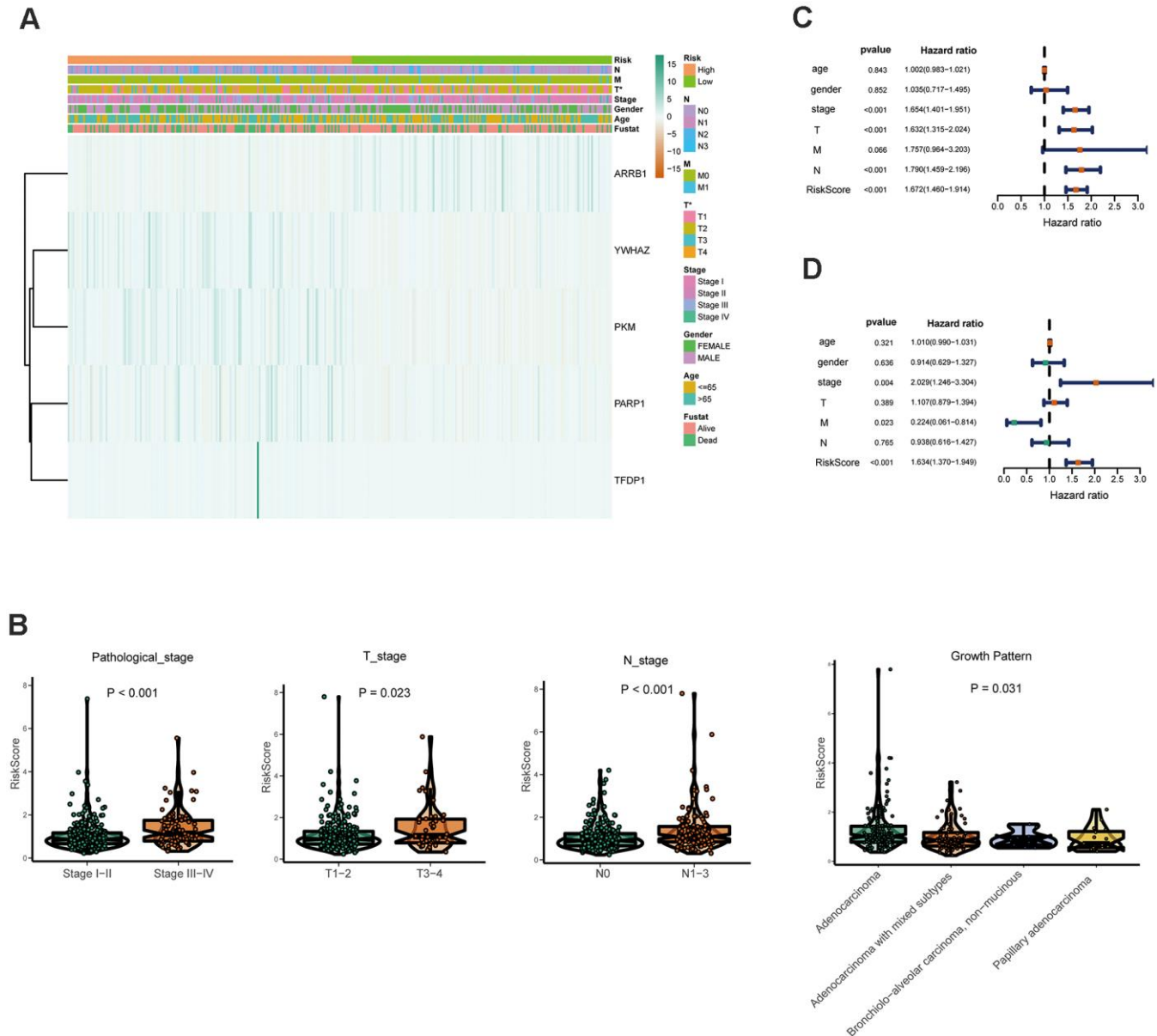
**Figure 5. Construction of the epigenetic-based prognostic risk signature in the TCGA cohort. (A)** The risk score distribution of LUAD patients. **(B)** Survival status and duration of patients. **(C)** Survival curves for the low risk and high risk groups. **(D)** Time-independent receiver operating characteristic (ROC) analysis of risk scores for prediction the overall survival in the TCGA set.

positively correlated with advanced tumor stage, higher T stage, and higher N stage ( $P < 0.05$ , Table 2).

### Validating the performance of the epigenetic-related signature

Both GSE31210 and GSE50081 datasets, including 174 and 127 LUAD samples, respectively, were used for validation. Consistent with our above results, the Kaplan-Meier curve showed that patients in the high risk group had a worse prognosis (Figures 7D, 8D). The risk score

(Figures 7A, 8A), survival status (Figures 7B, 8B) and gene expression heatmap (Figures 7C, 8C) of five ERGs are shown, respectively. The ROC curves were built to verify the predicted capability of five ERGs, and the AUC values for 1, 3 and 5-year survival were 0.777, 0.73, 0.746 (GSE31210, Figure 7E) and 0.711, 0.691, 0.735 (GSE50081, Figure 8E), respectively. A nomogram for predicting OS was established based on the five ERGs (Figure 9A). Moreover, the 1, 3, and 5-year OS predicted by our nomogram were remarkable consistent with the actual observed survival (Figure 9B–9D).



**Figure 6. Relationships between the risk assessment score expression and clinicopathological factors in LUAD.** (A) Heatmap of the epigenetic-related genes expression in the high and low risk groups and the clinicopathologic differences between the two groups. (B) Boxplots show the risk assessment score of patients with different pathological stage, T\_stage, N\_stage and growth pattern. Univariate (C) and multivariate (D) independent prognostic analysis of independent risk factors for overall survival (OS) in patients with LUAD.



**Table 2. Relationship between the expression of epigenetic-related genes and clinical characteristics.**

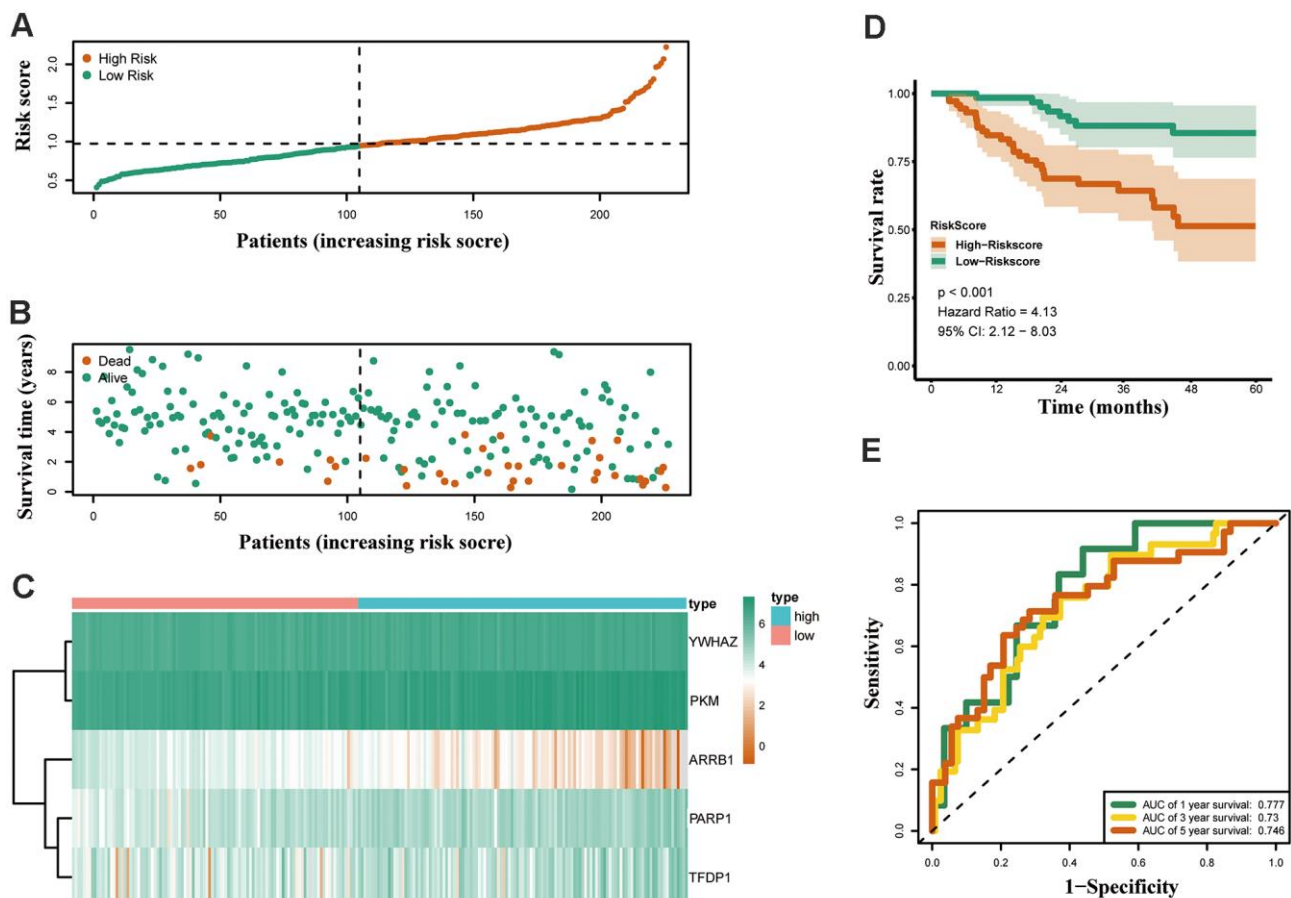
Gene symbol	Age (≥65/<65)	Gender (male/female)	Pathological stage (IV-III/ I-II)	T stage (T3-T4/T1-T2)	N stage (N2-3/N0-N1)	M stage (M1/ M0)
	t(P)	t(P)	t(P)	t(P)	t(P)	t(P)
ARRB1	-1.513(0.131)	-1.635(0.103)	1.487(0.139)	0.507(0.614)	1.17(0.243)	0.395(0.697)
PARP1	0.934(0.351)	-0.761(0.447)	-0.344(0.732)	-0.269(0.789)	-0.628(0.531)	-1.23(0.233)
YWHAZ	0.369(0.713)	-1.728(0.085)	<b>-2.189(0.031)</b>	-1.72(0.092)	-1.302(0.194)	-0.694(0.495)
TFDP1	1.113(0.267)	-1.077(0.283)	-1.035(0.304)	1.155(0.249)	-1.083(0.281)	-1.029(0.316)
PKM	-1.609(0.109)	-0.663(0.508)	<b>-2.865(0.005)</b>	-1.936(0.059)	<b>-3.006(0.003)</b>	-1.352(0.191)
riskScore	0.612(0.541)	-1.205(0.229)	<b>-3.031(0.003)</b>	<b>-2.092(0.042)</b>	<b>-2.704(0.008)</b>	-1.997(0.060)

Note: t: t value of student's t test; P: P-value of student's t test.

### Exploring the expression and prognostic of crucial ERGs

The mRNA expression of the five hub genes in LUAD were analyzed by using the TCGA database. We

confirmed that the mRNA expression of PARP1, PKM, TFDP1, and YWHAZ in LUAD tissues were all increased, while ARRB1 was decreased (Figure 10). Moreover, we explored the protein expression of the five hub genes. The results indicated that the protein of



**Figure 7. Validation of the epigenetic-based prognostic risk signature in the GSE31210 cohort.** (A) The risk score distribution of LUAD patients. (B) Survival status and duration of patients. (C) Heatmap of the epigenetic-related genes expression. (D) Survival curves for the low risk and high risk groups. (E) Time-independent receiver operating characteristic (ROC) analysis of risk scores for predicting the overall survival in the GSE31210 set.



PARP1, PKM, TFDP1, and YWHAZ were increased in LUAD tissues, which were in line with their mRNA expression levels (Figure 11). Furthermore, we found that the expression of the three high risk genes (PKM, TFDP1, and YWHAZ) were negatively associated with the prognosis in LUAD, while low risk gene ARRB1 was positive correlation with the prognosis by using the GEPIA database (Figure 12).

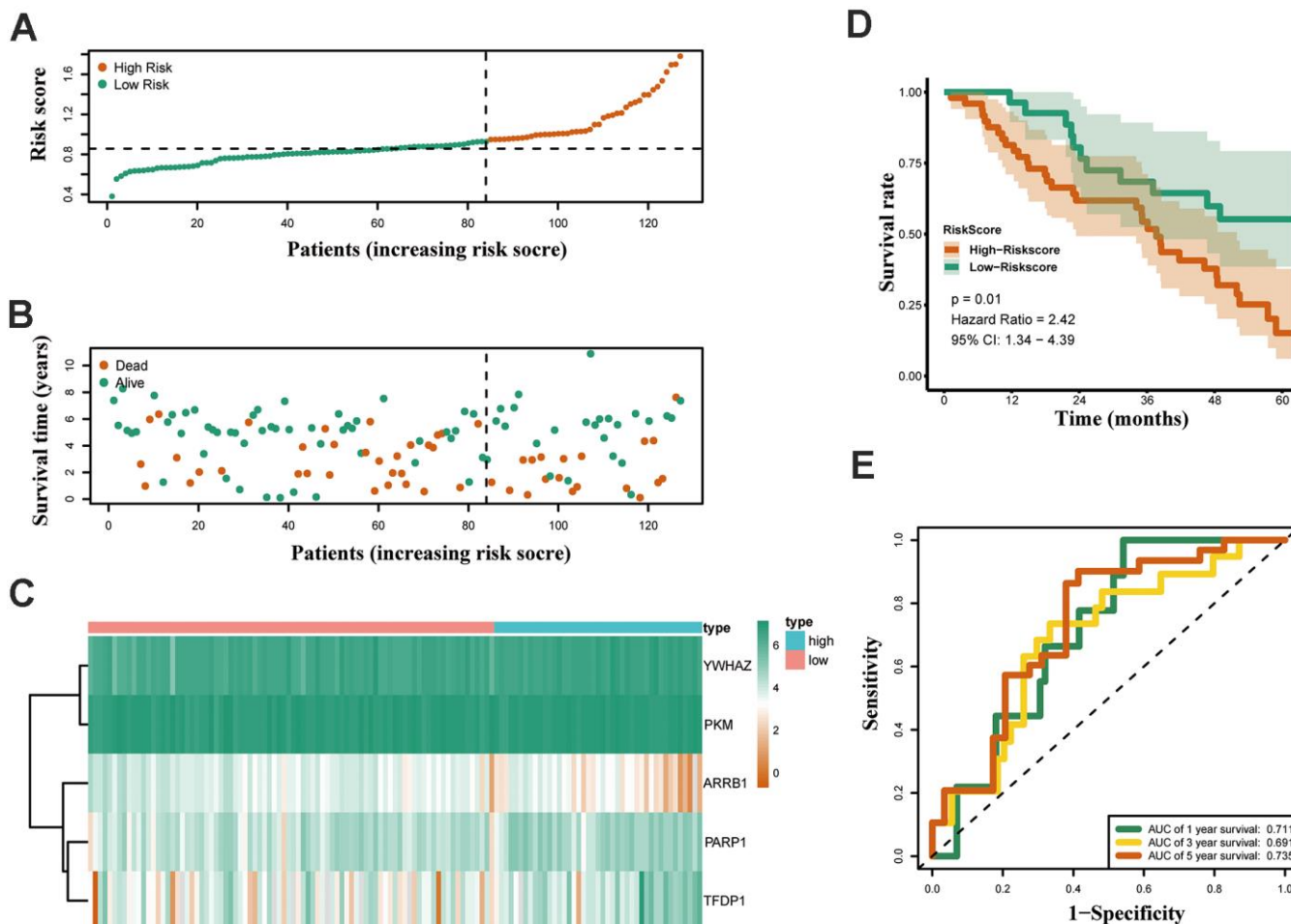
### Identification of potential candidate drugs targeting the signature

CMap mode-of-action (MoA) analysis found that 47 compounds could be candidate drugs targeting the signature and the drugs shared 18 kinds of mechanisms (Figure 13), ten drugs shared the MoA of topoisomerase inhibitor (amonafile, amsacrine, doxorubicin, ellipticine, etoposide, idarubicin, mitoxantrone, pidorubicine, pirarubicin, teniposide), seven drugs shared the MoA of

CDK inhibitor (aloesine, alvocidib, indirubin, indirubin, kenpaullone, purvalanol-a, roscovitine), six drugs shared the MoA of aurora kinase inhibitor (alisertib, barasertib, danusertib, methylnorlichexanthone, reversine, tozasertib), and five drugs shared the MoA of PARP inhibitor (olaparib, phenanthridone, rucaparib, veliparib, 3-amino-benzamide).

### DISCUSSION

Lung adenocarcinoma, a molecularly complex and heterogeneous disease, remains the most common causes of cancer-associated deaths worldwide [1, 20]. Recently, the development of drugs targeting on EGFR [5] and ALK [7] has promoted the treatment of some patients with LUAD, however, the highly gene mutation heterogeneous of LUAD has limited the benefits to a small number of patients. Notably, there is growing evidence that epigenetic modifications are frequently



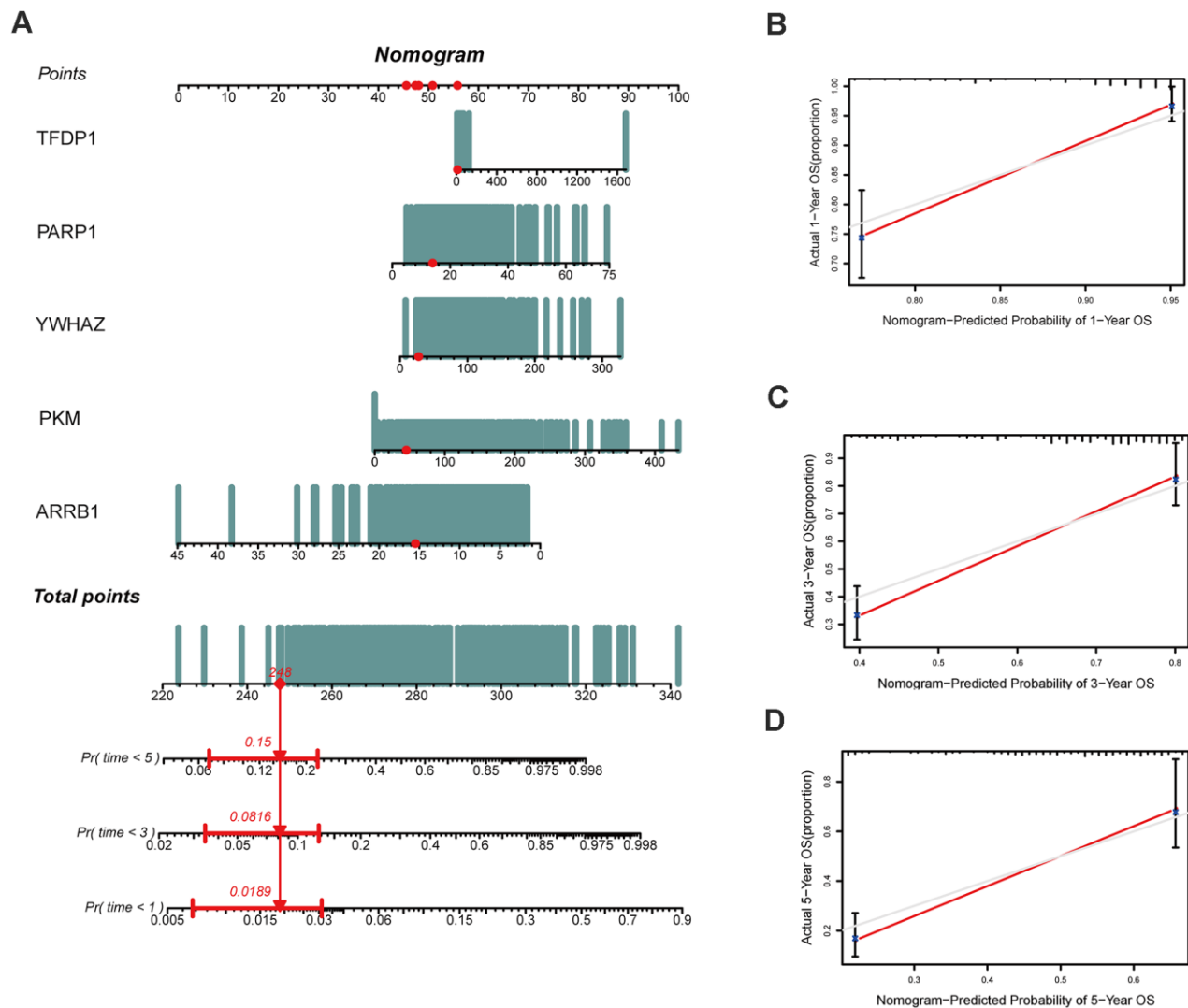
**Figure 8. Validation of the epigenetic-based prognostic risk signature in the GSE50081 cohort.** (A) The risk score distribution of LUAD patients. (B) Survival status and duration of patients. (C) Heatmap of the epigenetic-related genes expression. (D) Survival curves for the low risk and high risk groups. (E) Time-independent receiver operating characteristic (ROC) analysis of risk scores for predicting the overall survival in the GSE50081 set.

reversible, and may serve as attractive targets for lung cancer therapy [21, 22]. Therefore, it is particularly important to identify suitable epigenetic markers for LUAD diagnosis and prognosis, that can provide valuable support in decision making when considering treatment options.

In this study, we developed a novel and meaningful epigenetic-related prognostic signature (ARRB1, YWHAZ, PKM, PARP1, and TFDP1) for LUAD patients and validated them in two independent datasets from GEO. According our risk score system, LUAD patients were stratified into high and low risk groups, and patients in the high risk group had a worse prognosis. Then, we confirmed that each of the five signatures could be an independent prognostic factor. A new nomogram was constructed based on the five hub

genes, which can predict the OS of every LUAD patients. Interestingly, the expression of three high risk genes (PKM, TFDP1, and YWHAZ) had an enormous implication in the prognosis of LUAD patients. Finally, we identified 47 compounds that could serve as candidate targeted chemotherapeutic drugs to the five ERGs for LUAD patients. These drugs included the topoisomerase inhibitor, CDK inhibitor, aurora kinase inhibitor, PARP inhibitor, and so on.

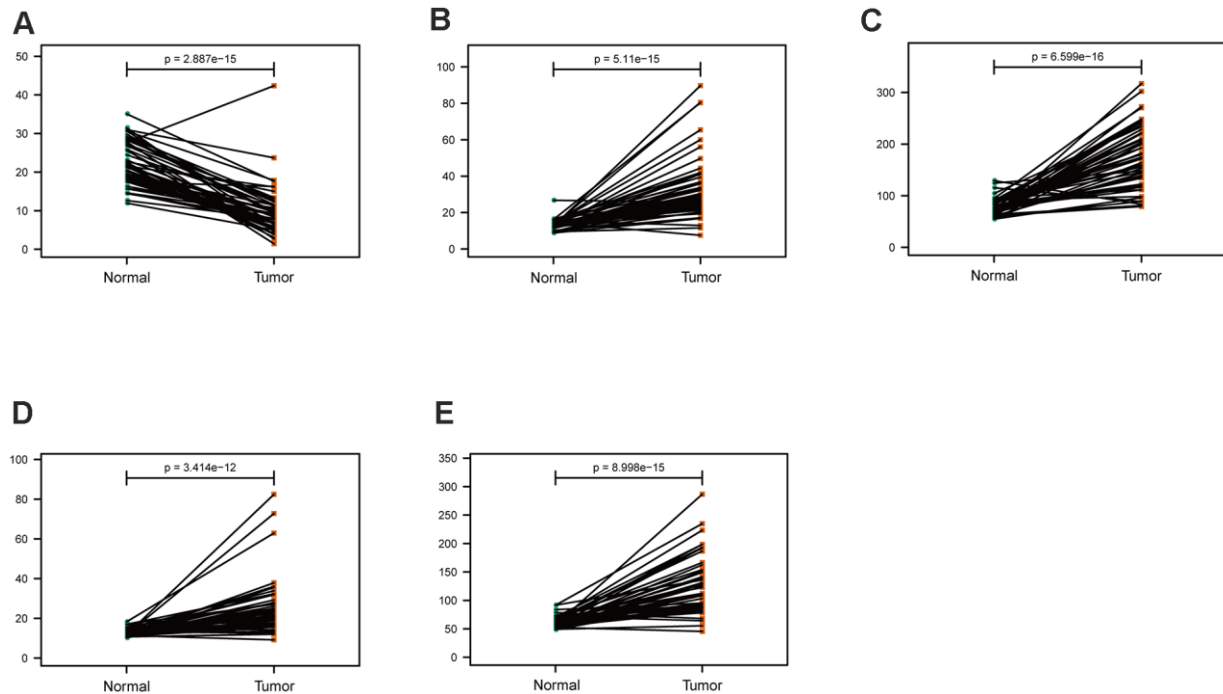
The epigenetic factors included in our signature have been previously demonstrated to be closely related to the development and progression of lung cancer. ARRB1 (also known as  $\beta$ -arrestin-1), a multifunctional adaptor, was initially discovered to promote internalization and desensitization of G protein-coupled receptors (GPCRs) [23, 24]. A study by Pillai et al. revealed that ARRB1



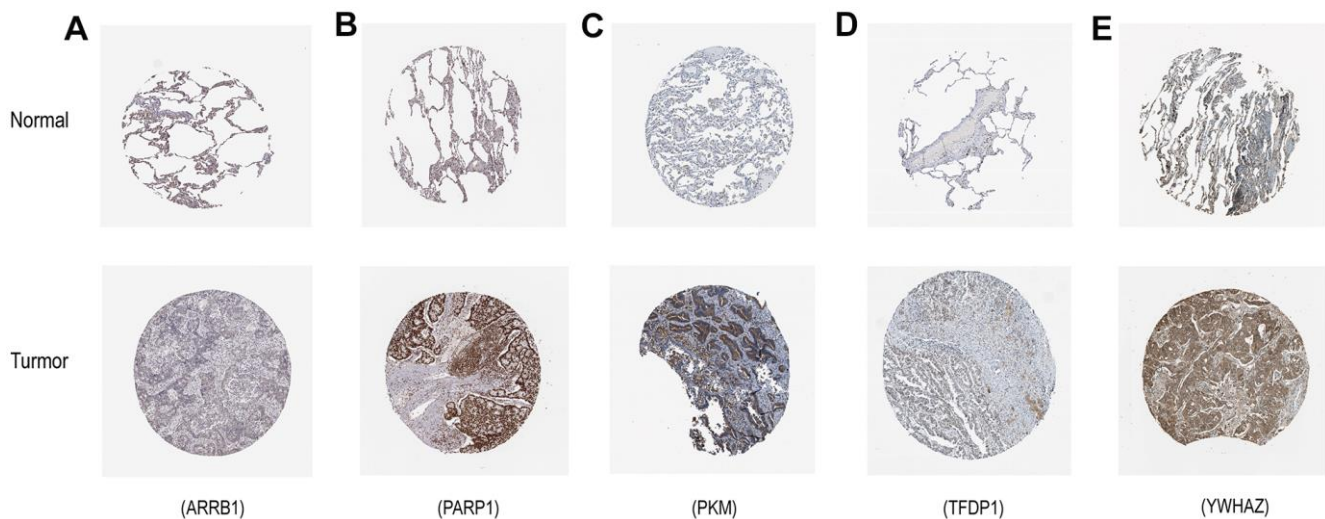
**Figure 9. Construction of a nomogram based on the epigenetic-related signature in the TCGA cohort. (A)** The nomogram based on the signature in LUAD patients at 1, 3, and 5 years. **(B–D)** Calibration curves of nomogram for the signature at 1, 3, and 5 years.

promoted the expression of mesenchymal genes through mediation of the E2F1 transcription factor in non-small cell lung carcinoma cell lines (NSCLC) [25]. Likewise, Shen et al. reported that ARRB1 could enhance the chemo-sensitivity of lung cancer through the mediation of DNA damage response [26]. Our results suggested that ARRB1 may be a putative tumor suppressor gene in

LUAD. The M2 isoform of pyruvate kinase (PKM2) (also named PKM, PK3, THBP1), an essential enzyme involved in glycolysis, is known to mediate the conversion of glucose to lactate in cancer cells under normoxic conditions [27, 28]. Jing Li reported that phosphorylation of PKM2 and inactivation of STAT3 inhibited lung cancer cell proliferation [29]. Inhibitors of



**Figure 10. Comparison of the hub genes mRNA levels in paired adjacent normal tissues and LUAD tissues from the TCGA cohort. (A) ARRB1, (B) PARP1, (C) PKM, (D) TFDP1, (E) YWHAZ.**



**Figure 11. Verification of hub ERGs expression in LUAD and normal lung tissue using the HPA database. (A) ARRB1, (B) PARP1, (C) PKM, (D) TFDP1, (E) YWHAZ.**

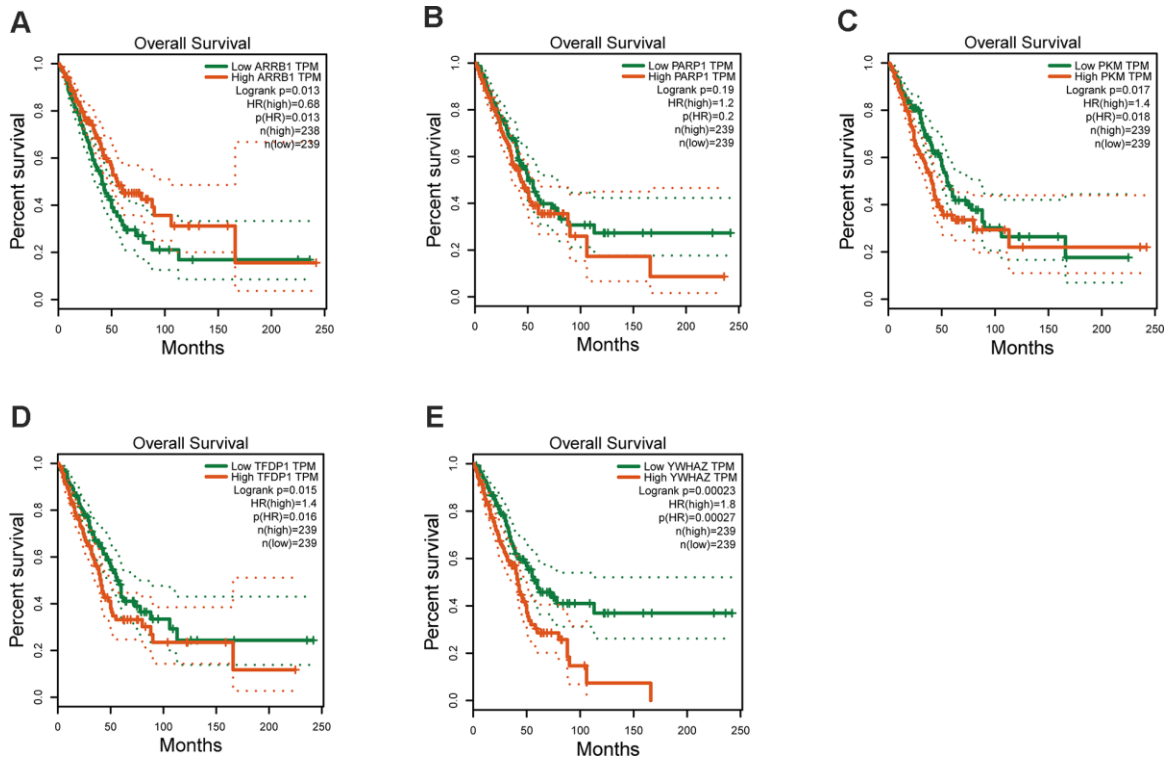


Figure 12. Kaplan–Meier survival analysis is performed on the hub genes. (A) ARRB1, (B) PARP1, (C) PKM, (D) TFDP1, (E) YWHAZ.

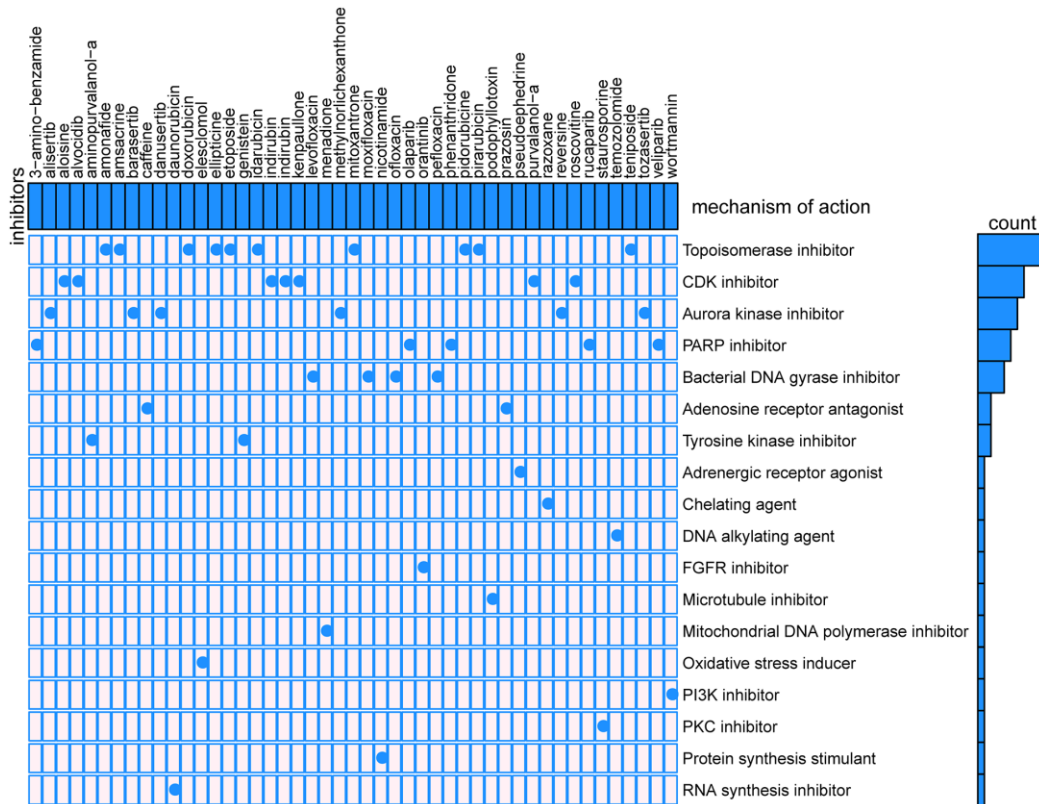


Figure 13. Results of CMap analysis.



PKM2 could moderately decelerate tumor cell proliferation [30, 31]. This finding might be in agreement with our current results, that PKM2 is a high risk gene in the context of LUAD. PARP1 has been reported as an abundant chromatin-associated enzyme able to catalyze the transfer of ADP-ribose units from NAD to substrate proteins [32]. Alternatively, PARP1 might play an important role in the development of various cancers, including cell proliferation, apoptosis, DNA repair, and so forth [33]. Recently, the use of immune checkpoint inhibitors has shown dramatic effect in the prognosis of NSCLC. Sophie Postel-Vinay reported that PARP inhibition enhanced the intrinsic immunity of tumor cells in NSCLC with deficiency in excision repair cross-complementing group 1, a gene which has crucial role in the nucleotide excision repair [34]. They also suggested that PARP inhibition, which did not cause severe bone marrow toxicity, might represent an interesting alternative or complement to platinum-based chemotherapy in combination with anti-PD-(L)1 agents in NSCLC. Transcription factor Dp-1 (TFDP1) is a key player of cell cycle regulation, it is a predominant protein that binds to E2F, another vital transcription protein that participates in the control of cell cycle and action of tumor suppressor proteins. TFDP1 can be candidate master regulators contributing to follicular lymphoma progression [35]. Knockdown of TFDP1 reduced both PITX1 promoter activity and mRNA transcription which caused patients suffering of knee/hip osteoarthritis [36]. Wang et al. revealed that TFDP1/E2F1 transcriptional activity played an important role in NSCLC [37]. YWHAZ (also named 14-3-3ζ), acts as a major hub protein for many signal transduction pathways [38, 39]. YWHAZ was frequently shown to be upregulated in several types of cancers, and its overexpression was often correlated with unfavorable prognosis of cancer patients [40–43]. Ma et al. found that YWHAZ was a credible prognostic biomarker, and might be a therapeutic target in NSCLC [44].

Overall, in concordance with our findings, there is mounting evidence in the literature that tells about the important role of the 5 hub genes (described above) in relation to NSCLC, which further supports the use of these hub genes as prognostic genes for LUAD patients. In recent years, a great deal of knowledge has been accumulated to help identify prognostic signature of different types of cancers, one of which is NSCLC. In our study, we not only identified the five powerful ERGs that are useful as prognostic signature, also provided potential drugs for targeted therapy. Interestingly, PARP inhibitors are currently regarded as a novel class of small molecule therapeutics for lung cancer. Henning Willers and colleagues found that PARP inhibitor by controlling ROS levels upon EGFR TKI treatment, promoted the sensitivity of EGFR-

mutated lung cancer to tyrosine kinase inhibitor treatment [45]. We hope our study will provide more choice for those patients not having EGFR mutation.

In the present study, there are some limitations that require mentioning. First, the three datasets are all retrospective and, hence, extrapolation based on these results may be difficult. To that end, the findings of this work should be validated, and further exploration using a larger multicenter prospective observational trial is desirable. Second, our findings have to be enriched by conducting long term *in vivo* experiments and additional *in vitro* experiments to investigate the functional role of the epigenetic factors associated with LUAD.

In summary, five ERGs (ARRB1, PARP1, PKM, TFDP1, and YWHAZ) are promising biomarkers for the diagnosis and prognosis of LUAD, which could provide valuable references to identify whether LUAD patients are at high risk. In addition, our findings may provide precise targeted chemotherapeutic drugs.

## MATERIALS AND METHODS

### Study subjects

The RNA-seq data of 497 LUAD and 54 normal lung samples were downloaded from The Cancer Genome Atlas (TCGA) database, and the RNA-seq data of 174 LUAD and 127 LUAD samples were downloaded from GSE31210 and GSE50081 dataset, respectively. A total of 720 epigenetic-related genes (ERGs) were retrieved from the EpiFactors database (<http://epifactors.autosome.ru/>).

### Identification and enrichment analysis of differentially expressed ERGs

Differentially expressed ERGs were obtained by using the Limma package in R software [46]. To further investigate the biological relevance of these genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted utilizing R “GOplot” package.

### Development of the epigenetic-related signature for patients with LUAD

We collected clinical information of LUAD patients who were followed for less than 2000 days in the TCGA database. Univariate Cox regression analysis was performed by R “survival” package to identify ERGs markedly related to OS. Genetic alterations in the candidate genes were obtained from cBioPortal (<https://www.cbioportal.org/>). Then, a novel epigenetic-related prognostic signature was developed by Lasso and multivariate Cox regression analyses [47, 48]. The

risk scores of LUAD patients were calculated according to the formula: the signature risk score =  $\Sigma (\beta_i \times \text{Expi})$ , where  $\beta_i$ , the coefficients, represented the weight of the respective signature and Expi represented the prognostic factors expression value as previously described [49]. According to the signature with identified prognostic factors, a nomogram for predicting the probability of OS was established.

### Validating the performance of the epigenetic-related signature

To validate the performance of the signature, the GSE31210 and GSE50081 datasets were considered as the validation cohort. The risk scores for LUAD patients were calculated by using the formula. Survival and ROC curve analyses were performed as described above.

### Exploring the expression and prognostic of crucial ERGs

To explore the prognostic value and the expression of these ERGs in LUAD, survival analysis was conducted on GEPIA database (<http://gepia.cancer-pku.cn>), the expression of these ERGs were confirmed on Human Protein Atlas (HPA) online database (<http://www.proteinatlas.org/>) [50].

### Predicting candidate small molecules for LUAD patients

To predict potential drugs for LUAD patients, we utilized the Broad Institute's Connectivity Map (CMap) to screen candidate molecule drugs as previously described [51, 49].

### Statistical analysis

In this study, all statistical analyses were conducted using Perl software (version 5.28.1) and R platform (version 3.6.1). The comparison of two paired groups was performed using the Wilcoxon rank-sum test. P value <0.05 was considered as statistically significant.

## AUTHOR CONTRIBUTIONS

Shenghui Qin conceived and directed the project, and wrote the manuscript. Zhihao Wang performed data bioinformatics analyses. Kidane Siele Embaye et al. helped with part of English writing and checking. All authors read and approved the manuscript.

## ACKNOWLEDGMENTS

All data generated or analyzed during the present study was downloaded from TCGA database, GEO database,

GEPIA database, and HPA database. We thank the authors who provided the data for this study.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This research was supported by a Grant from the National Natural Science Foundation of China (No. 81700012).

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019; 69:7–34. <https://doi.org/10.3322/caac.21551> PMID:30620402
2. Houston KA, Henley SJ, Li J, White MC, Richards TB. Patterns in lung cancer incidence rates and trends by histologic type in the United States, 2004-2009. *Lung Cancer.* 2014; 86:22–28. <https://doi.org/10.1016/j.lungcan.2014.08.001> PMID:25172266
3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015; 136:E359–86. <https://doi.org/10.1002/ijc.29210> PMID:25220842
4. Dolly SO, Collins DC, Sundar R, Popat S, Yap TA. Advances in the Development of Molecularly Targeted Agents in Non-Small-Cell Lung Cancer. *Drugs.* 2017; 77:813–27. <https://doi.org/10.1007/s40265-017-0732-2> PMID:28378229
5. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science.* 2004; 304:1497–500. <https://doi.org/10.1126/science.1099314> PMID:15118125
6. Muller IB, de Langen AJ, Giovannetti E, Peters GJ. Anaplastic lymphoma kinase inhibition in metastatic non-small cell lung cancer: clinical impact of alectinib. *Onco Targets Ther.* 2017; 10:4535–41. <https://doi.org/10.2147/OTT.S109493> PMID:28979145
7. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Jänne PA, Costa DB, Varella-Garcia M, Kim WH, Lynch TJ, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med.* 2010; 363:1693–703.

- <https://doi.org/10.1056/NEJMoa1006448>  
PMID:20979469
8. Grewal SI, Moazed D. Heterochromatin and epigenetic control of gene expression. *Science*. 2003; 301:798–802.  
<https://doi.org/10.1126/science.1086887>  
PMID:12907790
9. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet*. 2002; 3:662–73.  
<https://doi.org/10.1038/nrg887> PMID:12209141
10. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science*. 2017; 357:eaal2380.  
<https://doi.org/10.1126/science.aal2380>  
PMID:28729483
11. Suvà ML, Riggi N, Bernstein BE. Epigenetic reprogramming in cancer. *Science*. 2013; 339:1567–70.  
<https://doi.org/10.1126/science.1230184>  
PMID:23539597
12. Yuan S, Yu Z, Liu Q, Zhang M, Xiang Y, Wu N, Wu L, Hu Z, Xu B, Cai T, Ma X, Zhang Y, Liao C, et al. GPC5, a novel epigenetically silenced tumor suppressor, inhibits tumor growth by suppressing Wnt/ $\beta$ -catenin signaling in lung adenocarcinoma. *Oncogene*. 2016; 35:6120–31.  
<https://doi.org/10.1038/onc.2016.149> PMID:27157618
13. Capelletto E, Novello S. Emerging new agents for the management of patients with non-small cell lung cancer. *Drugs*. 2012 (Suppl 1); 72:37–52.  
<https://doi.org/10.2165/1163028-S0-000000000-00000> PMID:22712796
14. Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B, Sebree R, Rodgers K, Hooker CM, Franco N, Lee B, Tsai S, Delgado IE, et al. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. *Cancer Discov*. 2011; 1:598–607.  
<https://doi.org/10.1158/2159-8290.CD-11-0214>  
PMID:22586682
15. Lok BH, Gardner EE, Schneeberger VE, Ni A, Desmeules P, Rekhman N, de Stanchina E, Teicher BA, Riaz N, Powell SN, Poirier JT, Rudin CM. PARP Inhibitor Activity Correlates with SLFN11 Expression and Demonstrates Synergy with Temozolomide in Small Cell Lung Cancer. *Clin Cancer Res*. 2017; 23:523–35.  
<https://doi.org/10.1158/1078-0432.CCR-16-1040>  
PMID:27440269
16. Li W, Gao LN, Song PP, You CG. Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging (Albany NY)*. 2020; 12:3558–73.  
<https://doi.org/10.18632/aging.102828>  
PMID:32087603
17. Jiang H, Xu S, Chen C. A ten-gene signature-based risk assessment model predicts the prognosis of lung adenocarcinoma. *BMC Cancer*. 2020; 20:782.  
<https://doi.org/10.1186/s12885-020-07235-z>  
PMID:32819300
18. Liu Y, Guo X, Zhao M, Ao H, Leng X, Liu M, Wu C, Ma J, Zhu J. Contributions and prognostic values of m<sup>6</sup>A RNA methylation regulators in non-small-cell lung cancer. *J Cell Physiol*. 2020; 235:6043–57.  
<https://doi.org/10.1002/jcp.29531> PMID:32052446
19. Luo WM, Wang ZY, Zhang X. Identification of four differentially methylated genes as prognostic signatures for stage I lung adenocarcinoma. *Cancer Cell Int*. 2018; 18:60.  
<https://doi.org/10.1186/s12935-018-0547-6>  
PMID:29713243
20. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med*. 2008; 359:1367–80.  
<https://doi.org/10.1056/NEJMra0802714>  
PMID:18815398
21. Liu SV, Fabbri M, Gitlitz BJ, Laird-Offringa IA. Epigenetic therapy in lung cancer. *Front Oncol*. 2013; 3:135.  
<https://doi.org/10.3389/fonc.2013.00135>  
PMID:23755372
22. Mehta A, Dobersch S, Romero-Olmedo AJ, Barreto G. Epigenetics in lung cancer diagnosis and therapy. *Cancer Metastasis Rev*. 2015; 34:229–41.  
<https://doi.org/10.1007/s10555-015-9563-3>  
PMID:25939322
23. Lefkowitz RJ, Whalen EJ. beta-arrestins: traffic cops of cell signaling. *Curr Opin Cell Biol*. 2004; 16:162–68.  
<https://doi.org/10.1016/j.ceb.2004.01.001>  
PMID:15196559
24. Lefkowitz RJ, Shenoy SK. Transduction of receptor signals by beta-arrestins. *Science*. 2005; 308:512–17.  
<https://doi.org/10.1126/science.1109237>  
PMID:15845844
25. Pillai S, Trevino J, Rawal B, Singh S, Kovacs M, Li X, Schell M, Haura E, Bepler G, Chellappan S.  $\beta$ -arrestin-1 mediates nicotine-induced metastasis through E2F1 target genes that modulate epithelial-mesenchymal transition. *Cancer Res*. 2015; 75:1009–20.  
<https://doi.org/10.1158/0008-5472.CAN-14-0681>  
PMID:25600647
26. Shen H, Wang L, Zhang J, Dong W, Zhang T, Ni Y, Cao H, Wang K, Li Y, Wang Y, Du J. ARRB1 enhances the chemosensitivity of lung cancer through the mediation of DNA damage response. *Oncol Rep*. 2017; 37:761–67.

- <https://doi.org/10.3892/or.2016.5337>  
PMID:[28035404](https://pubmed.ncbi.nlm.nih.gov/28035404/)
27. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144:646–74.  
<https://doi.org/10.1016/j.cell.2011.02.013>  
PMID:[21376230](https://pubmed.ncbi.nlm.nih.gov/21376230/)
28. Wong N, Ojo D, Yan J, Tang D. PKM2 contributes to cancer metabolism. *Cancer Lett*. 2015; 356:184–91.  
<https://doi.org/10.1016/j.canlet.2014.01.031>  
PMID:[24508027](https://pubmed.ncbi.nlm.nih.gov/24508027/)
29. Xiao M, Xie J, Wu Y, Wang G, Qi X, Liu Z, Wang Y, Wang X, Hoque A, Oakhill J, Proud CG, Li J. The eEF2 kinase-induced STAT3 inactivation inhibits lung cancer cell proliferation by phosphorylation of PKM2. *Cell Commun Signal*. 2020; 18:25.  
<https://doi.org/10.1186/s12964-020-0528-y>  
PMID:[32054489](https://pubmed.ncbi.nlm.nih.gov/32054489/)
30. Spoden GA, Mazurek S, Morandell D, Bacher N, Ausserlechner MJ, Jansen-Dürr P, Eigenbrodt E, Zwerschke W. Isozyme-specific inhibitors of the glycolytic key regulator pyruvate kinase subtype M2 moderately decelerate tumor cell proliferation. *Int J Cancer*. 2008; 123:312–21.  
<https://doi.org/10.1002/ijc.23512> PMID:[18425820](https://pubmed.ncbi.nlm.nih.gov/18425820/)
31. Li RZ, Fan XX, Shi DF, Zhu GY, Wang YW, Luo LX, Pan HD, Yao XJ, Leung EL, Liu L. Identification of a new pyruvate kinase M2 isoform (PKM2) activator for the treatment of non-small-cell lung cancer (NSCLC). *Chem Biol Drug Des*. 2018; 92:1851–58.  
<https://doi.org/10.1111/cbdd.13354> PMID:[29931766](https://pubmed.ncbi.nlm.nih.gov/29931766/)
32. Amé JC, Spenlehauer C, de Murcia G. The PARP superfamily. *Bioessays*. 2004; 26:882–93.  
<https://doi.org/10.1002/bies.20085> PMID:[15273990](https://pubmed.ncbi.nlm.nih.gov/15273990/)
33. Gibson BA, Kraus WL. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. *Nat Rev Mol Cell Biol*. 2012; 13:411–24.  
<https://doi.org/10.1038/nrm3376> PMID:[22713970](https://pubmed.ncbi.nlm.nih.gov/22713970/)
34. Chabanon RM, Muirhead G, Krastev DB, Adam J, Morel D, Garrido M, Lamb A, Hénon C, Dorvault N, Rouanne M, Marlow R, Bajrami I, Cardeñosa ML, et al. PARP inhibition enhances tumor cell-intrinsic immunity in ERCC1-deficient non-small cell lung cancer. *J Clin Invest*. 2019; 129:1211–28.  
<https://doi.org/10.1172/JCI123319>  
PMID:[30589644](https://pubmed.ncbi.nlm.nih.gov/30589644/)
35. Bisikirska B, Bansal M, Shen Y, Teruya-Feldstein J, Chaganti R, Califano A. Elucidation and Pharmacological Targeting of Novel Molecular Drivers of Follicular Lymphoma Progression. *Cancer Res*. 2016; 76:664–74.  
<https://doi.org/10.1158/0008-5472.CAN-15-0828>  
PMID:[26589882](https://pubmed.ncbi.nlm.nih.gov/26589882/)
36. Pellicelli M, Picard C, Wang D, Lavigne P, Moreau A. E2F1 and TFDP1 Regulate PITX1 Expression in Normal and Osteoarthritic Articular Chondrocytes. *PLoS One*. 2016; 11:e0165951.  
<https://doi.org/10.1371/journal.pone.0165951>  
PMID:[27802335](https://pubmed.ncbi.nlm.nih.gov/27802335/)
37. Zhan W, Wang W, Han T, Xie C, Zhang T, Gan M, Wang JB. COMMD9 promotes TFDP1/E2F1 transcriptional activity via interaction with TFDP1 in non-small cell lung cancer. *Cell Signal*. 2017; 30:59–66.  
<https://doi.org/10.1016/j.cellsig.2016.11.016>  
PMID:[27871936](https://pubmed.ncbi.nlm.nih.gov/27871936/)
38. Aitken A. Post-translational modification of 14-3-3 isoforms and regulation of cellular function. *Semin Cell Dev Biol*. 2011; 22:673–80.  
<https://doi.org/10.1016/j.semcdb.2011.08.003>  
PMID:[21864699](https://pubmed.ncbi.nlm.nih.gov/21864699/)
39. Yu CC, Li CF, Chen IH, Lai MT, Lin ZJ, Korla PK, Chai CY, Ko G, Chen CM, Hwang T, Lee SC, Sheu JJ. YWHAZ amplification/overexpression defines aggressive bladder cancer and contributes to chemo-/radio-resistance by suppressing caspase-mediated apoptosis. *J Pathol*. 2019; 248:476–87.  
<https://doi.org/10.1002/path.5274> PMID:[30945298](https://pubmed.ncbi.nlm.nih.gov/30945298/)
40. Li Y, Wang J, Dai X, Zhou Z, Liu J, Zhang Y, Li Y, Hou Y, Pang L, Wang X, Wang C, Hao Z, Zhang Y, et al. miR-451 regulates FoxO3 nuclear accumulation through Ywhaz in human colorectal cancer. *Am J Transl Res*. 2015; 7:2775–85.  
PMID:[26885274](https://pubmed.ncbi.nlm.nih.gov/26885274/)
41. Tong S, Xia T, Fan K, Jiang K, Zhai W, Li JS, Wang SH, Wang JJ. 14-3-3 $\zeta$  promotes lung cancer cell invasion by increasing the Snail protein expression through atypical protein kinase C (aPKC)/NF- $\kappa$ B signaling. *Exp Cell Res*. 2016; 348:1–9.  
<https://doi.org/10.1016/j.yexcr.2016.08.014>  
PMID:[27554601](https://pubmed.ncbi.nlm.nih.gov/27554601/)
42. Wang W, Zhang L, Wang Y, Ding Y, Chen T, Wang Y, Wang H, Li Y, Duan K, Chen S, Yang Q, Chen C. Involvement of miR-451 in resistance to paclitaxel by regulating YWHAZ in breast cancer. *Cell Death Dis*. 2017; 8:e3071.  
<https://doi.org/10.1038/cddis.2017.460>  
PMID:[28981108](https://pubmed.ncbi.nlm.nih.gov/28981108/)
43. Gan Y, Ye F, He XX. The role of YWHAZ in cancer: A maze of opportunities and challenges. *J Cancer*. 2020; 11:2252–64.  
<https://doi.org/10.7150/jca.41316> PMID:[32127952](https://pubmed.ncbi.nlm.nih.gov/32127952/)
44. Deng Y, Zheng J, Ma J. The clinical and prognostic significance of YWHAZ in non-small-cell lung cancer patients: Immunohistochemical analysis. *J Cell Biochem*. 2019; 120:6290–98.



- <https://doi.org/10.1002/jcb.27915>  
PMID:[30378158](https://pubmed.ncbi.nlm.nih.gov/30378158/)
45. Marcar L, Bardhan K, Gheorghiu L, Dinkelborg P, Pfäffle H, Liu Q, Wang M, Piotrowska Z, Sequist LV, Borgmann K, Settleman JE, Engelman JA, Hata AN, Willers H. Acquired Resistance of EGFR-Mutated Lung Cancer to Tyrosine Kinase Inhibitor Treatment Promotes PARP Inhibitor Sensitivity. *Cell Rep.* 2019; 27:3422–3432.e4.  
<https://doi.org/10.1016/j.celrep.2019.05.058>  
PMID:[31216465](https://pubmed.ncbi.nlm.nih.gov/31216465/)
46. Noble WS. How does multiple testing correction work? *Nat Biotechnol.* 2009; 27:1135–37.  
<https://doi.org/10.1038/nbt1209-1135>  
PMID:[20010596](https://pubmed.ncbi.nlm.nih.gov/20010596/)
47. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010; 33:1–22.  
PMID:[20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
48. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CG. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med.* 2018; 6:121.  
<https://doi.org/10.21037/atm.2018.02.12>  
PMID:[29955581](https://pubmed.ncbi.nlm.nih.gov/29955581/)
49. Wang Z, Embaye KS, Yang Q, Qin L, Zhang C, Liu L, Zhan X, Zhang F, Wang X, Qin S. A Novel Metabolism-Related Signature as a Candidate Prognostic Biomarker for Hepatocellular Carcinoma. *J Hepatocell Carcinoma.* 2021; 8:119–32.  
<https://doi.org/10.2147/JHC.S294108>  
PMID:[33758763](https://pubmed.ncbi.nlm.nih.gov/33758763/)
50. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, et al. A subcellular map of the human proteome. *Science.* 2017; 356:eaal3321.  
<https://doi.org/10.1126/science.aal3321>  
PMID:[28495876](https://pubmed.ncbi.nlm.nih.gov/28495876/)
51. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science.* 2006; 313:1929–35.  
<https://doi.org/10.1126/science.1132939>  
PMID:[17008526](https://pubmed.ncbi.nlm.nih.gov/17008526/)