



OPEN ACCESS

EDITED BY
Bin Liu,
Jiangsu Ocean University, China

REVIEWED BY
Jing Ji,
Jiangsu Ocean University, China
An Qin,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Linhua Liu,
liulinhua@hotmail.com

[†]These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Genetics

RECEIVED 01 August 2022
ACCEPTED 08 September 2022
PUBLISHED 23 September 2022

CITATION
Huang S, Chen S, Zhang D, Gao J and
Liu L (2022), Enhancer-associated
regulatory network and gene signature
based on transcriptome and
methylation data to predict the survival
of patients with lung adenocarcinoma.
Front. Genet. 13:1008602.
doi: 10.3389/fgene.2022.1008602

COPYRIGHT
© 2022 Huang, Chen, Zhang, Gao and
Liu. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Enhancer-associated regulatory network and gene signature based on transcriptome and methylation data to predict the survival of patients with lung adenocarcinoma

Shihao Huang^{1†}, Shiyu Chen^{2†}, Di Zhang¹, Jiamei Gao¹ and Linhua Liu^{1*}

¹Department of Biochemistry, Institute of Glycobiology, Dalian Medical University, Dalian, Liaoning, China, ²Department of Laboratory Medicine, Nanxishan Hospital of Guangxi Zhuang Autonomous Region, Guilin, China

Accumulating evidence has proved that aberrant methylation of enhancers plays regulatory roles in gene expression for various cancers including lung adenocarcinoma (LUAD). In this study, the transcriptome and methylation data of The Cancer Genome Atlas (TCGA)-LUAD cohort were comprehensively analyzed with a five-step Enhancer Linking by Methylation/Expression Relationships (ELMER) process. Step 1: 131,371 distal (2 kb upstream from the transcription start site) probes were obtained. Step 2: 10,665 distal hypomethylated probes were identified in an unsupervised mode with the `get.diff.meth` function. Step 3: 699 probe-gene pairs with negative correlations were screened using the `get.pair` function in an unsupervised mode. Step 4: After mapping with probes, 768 motifs were obtained and 24 of them were enriched. Step 5: 127 transcription factors (TFs) with differential expressions and negative correlations with methylation levels were screened, which were corresponding to 21 motifs. After the ELMER process, a prognostic “TFs-motifs-genes” regulatory network was constructed. The Least absolute shrinkage and selection operator (LASSO) and Stepwise regression analyses were further applied to identify variables in the TCGA-LUAD cohort and an eight-gene signature was constructed for calculating the risk score. The risk score was verified in two independent validation cohorts. The area under curve values of receiver operating characteristic curves predicting 1-, 3-, and 5-years survival ranged from 0.633 to 0.764. With the increase of the risk scores, both the survival statuses and clinical traits showed a worse tendency. There were significant differences in the degrees of immune cell infiltration, TMB values, and TIDE scores between the high-risk and low-risk groups. Finally, a better-performing prognostic nomogram was integrated with the risk score and other clinical traits. In short, this multi-omics analysis demonstrated the application of ELMER in analyzing enhancer-associated regulatory network in

LUAD, which provided promising strategies for epigenetic therapy and prognostic biomarkers.

KEYWORDS

lung adenocarcinoma, enhancer, methylation, regulatory network, gene signature, prognostic model

Introduction

Enhancer is a DNA sequence in the genome with a length of 50–1,500 bp, which can bind with transcription factors (TFs) to promote the transcription of the target gene. The position of enhancer is not fixed and can be at the near end or the far end of a target gene. The enhancer may be upstream or downstream of its regulatory gene (Field and Adelman, 2020).

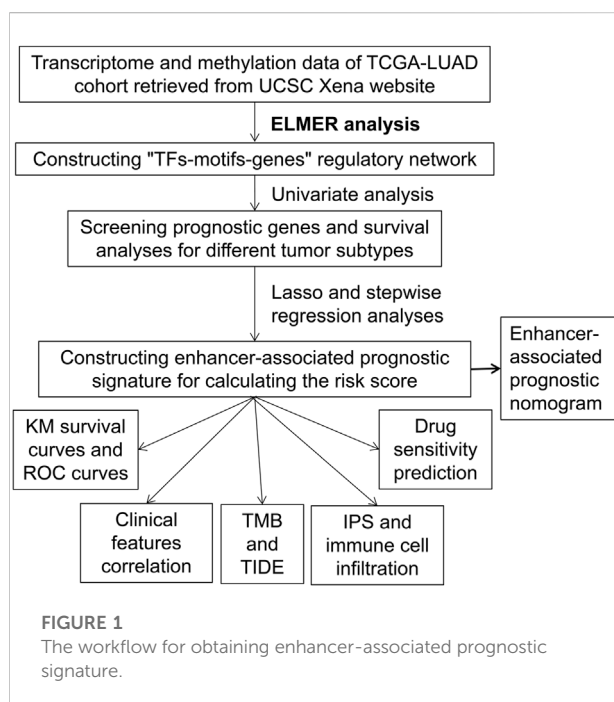
Enhancers have been reported to reflect normal and pathogenic cellular conditions (Fishilevich et al., 2017). Some high-throughput identification approaches have been developed to predict the enhancers and corresponding functions (Kleftogiannis et al., 2016). With the progress of functions, enhancers have been found to link with several diseases (Wang et al., 2018). Researchers have tried to build a database for disease-associated enhancers: DiseaseEnhancer (Zhang et al., 2018).

Reviews are describing the roles of enhancers in tumors (Sur and Taipale, 2016). As important regulatory elements of DNA, enhancers participate in several comprehensive regulatory networks of cancer-associated genes. Mutations in tumors often lead to aberrantly regulated enhancers, as well as abnormal expression of growth-related genes (Adhikary et al.,

2021). The abnormal regulation can be *trans*-action, such as the activation of a transcription factor or apparent regulatory factors that control enhancer activity. Similarly, abnormal regulation can also be *cis*-action, such as mutation to change enhancer activity or its specificity to the target gene (Chen and Liang, 2020). Investigating the activity regulation and related mechanism of tumor type-specific enhancers at the molecular level may be applied for screening therapeutic targets.

Accumulating evidence has suggested that many aberrant methylation sites have been observed on enhancer sequences in cancer cells (Herz, 2016). These abnormal methylations have been proven to link with the expression of a target gene, as well as the disease progression. The methylation state of enhancer regions is the promising next generation of epigenetic biomarkers (Clermont et al., 2016). One study has revealed the abnormal enhancer of hepatocellular carcinoma (HCC) based on multi-omics data (Xiong et al., 2019). By comprehensive analysis of ChIP-seq data, transcriptome data, DNA methylation data, and HiC data, the abnormal enhancer and related transcription disorders in HCC have been described, and the differentially methylated enhancer and its target genes were identified. A prognostic model based on these differentially expressed genes (DEGs) of abnormal enhancers was constructed, which predicted the prognosis of HCC. It was beneficial to the development of epigenetic therapy for HCC (Huang et al., 2022). However, the epigenetic regulation and function of transcription enhancers have been still unclear.

Lung cancer is the leading cause of cancer mortality. It is classified into various histologic subtypes, including adenocarcinoma, squamous carcinoma, non-small cell lung cancer, and small cell lung cancer (Ruiz-Cordero and Devine, 2020). Lung cancer exhibited a good response to novel targeted therapies, such as checkpoint immunotherapy. With the advances in knowledge on molecular characteristics of lung cancer, researchers have found different treatment decisions should be provided to patients with varied gene expression profiles (Chen et al., 2020). For example, some patients may show a better response to immunotherapy, while others should receive targeted therapies and chemotherapy before considering immunotherapy as a single agent (Mazieres et al., 2019). Drug resistance should also be considered (Denisenko et al., 2018). Some studies have tried to find the diagnostic and prognostic prediction markers for lung cancers, such as the immune-related genes signature (Yi et al., 2021), m6A modification (Li et al., 2021), and costimulatory molecule-based signature (Zhang et al.,



2020). All these efforts have contributed to the precision and individual treatment of patients with tumors.

This study aims to find enhancer-associated prognostic biomarkers. The transcriptome and methylation of lung adenocarcinoma (LUAD) cohorts have been integrated. The enhancer-associated regulatory network was constructed after Enhancer Linking by Methylation/Expression Relationships (ELMER) analysis. An enhancer-associated prognostic gene signature has been constructed with the screened target transcription factor and target genes, exhibiting good prognostic prediction performance for patients with lung cancers. The complete analysis route of this study has been provided in Figure 1.

Materials and methods

Datasets

Three cohorts were applied: The Cancer Genome Atlas (TCGA)-LUAD, GSE31210, and GSE8894. Both the transcriptome and methylation data of TCGA-LUAD were retrieved from the UCSC Xena website, and the transcriptome data of GSE31210 and GSE8894 were obtained from the Gene Expression Omnibus (GEO) database. The clinical features were also obtained, and the data from patients with overall survival of <30 days were removed. All transcriptomic data were normalized with $\log_2(x+1)$ method and the combat function of the sva R package was used to exclude batch effects (Leek et al., 2012). The methylation data were normalized with the champ.norm function of the ChAMP R package, and the missing values in methylation data have been filled with function impute.knn (Tian et al., 2017).

Enhancer-associated regulatory network

With the gene methylation and transcription levels of the same batch of samples in the TCGA-LUAD cohort, the sequences of the differential methylation probes were screened with the ELMER R package (Silva et al., 2019), found the enriched motifs, and further predicted the TFs interacting with these motifs. Finally, the “TFs-motifs-genes” regulatory network was constructed. ELMER’s main analysis process consisted of the following five parts: 1) Identification of distal probes (probes larger than 2 kb upstream from the transcription start site) from methylation chip data; 2) identification of differences in methylation levels between normal and tumor groups; 3) identification of target genes for differentially methylated probes; 4) identification of motifs enriched with both differentially methylated and target gene-related probes; 5) identification of TFs based on transcriptional differences.

Identification of prognostic regulatory network and tumor subtypes

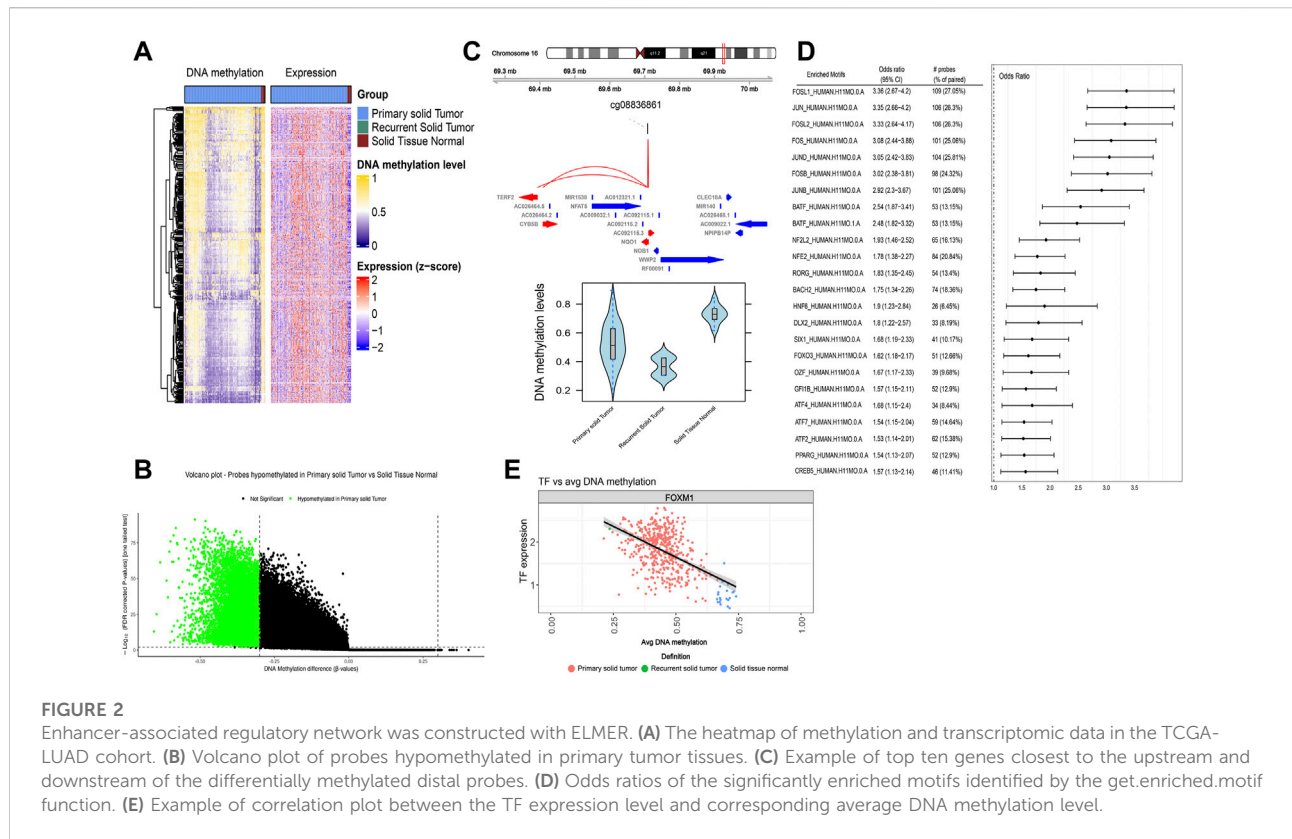
Based on the transcriptional level of genes included in the regulatory network, we first conducted the univariate analysis to screen prognostic genes, and then performed the unsupervised clustering in the three cohorts with the ConsensusClusterPlus R package by the k-means method (Wilkerson and Hayes, 2010). The clustering process was carried out 1,000 times, involving 80% samples in each iteration. Subsequently, survival analyses were performed for different subtypes.

Construction of enhancer-associated prognostic signature

TCGA-LUAD was applied as the training set, GSE31210 and GSE8894 were applied as the validation sets. Least absolute shrinkage and selection operator (LASSO) and Stepwise regression analyses were applied to further streamline prognostic variables in the TCGA-LUAD cohort and construct a multi-gene COX signature for calculating the risk score of each patient (Tibshirani, 1997). Patients were divided into high-risk and low-risk groups based on the median risk score. Then Kaplan–Meier (KM) survival curves and Receiver Operating Characteristic (ROC) curves were plotted to assess the prediction effect of the model. Independent prognostic analyses were applied for validating the independence of the risk score compared with other clinical features in the three cohorts. The Wilcoxon rank sum test was applied to evaluate correlations between risk score and clinical features in TCGA-LUAD cohort.

Analysis of tumor mutation burden (TMB), tumor immune dysfunction and exclusion (TIDE), and immune micro-environment.

The mutation data of TCGA-LUAD based on VarScan2 (Koboldt et al., 2012) were obtained from TCGA Database. Non-synonymous mutations were calculated and the TMB scores were obtained with the number of variants/the length of exons. Wilcoxon test was used to analyze the difference in TMB values. The Maftools R package was used to calculate and plot the somatic alterations landscapes (Mayakonda et al., 2018). The TIDE scores of the TCGA-LUAD cohort were obtained from TIDE (Jiang et al., 2018). The differences in the immune micro-environment were compared with the proportions of 22 types of immune cells estimated by Cibersort (Chen et al., 2018). The Cibersort R package was applied. The simulation was conducted 1,000 times with the parameter of perm = 1,000, QN = True. The samples with $p > 0.05$ were rejected and removed. The correlation between risk scores and various immune cell infiltration was further analyzed with the Spearman correlation test.



Analysis of tumor immunogenicity and drug susceptibility

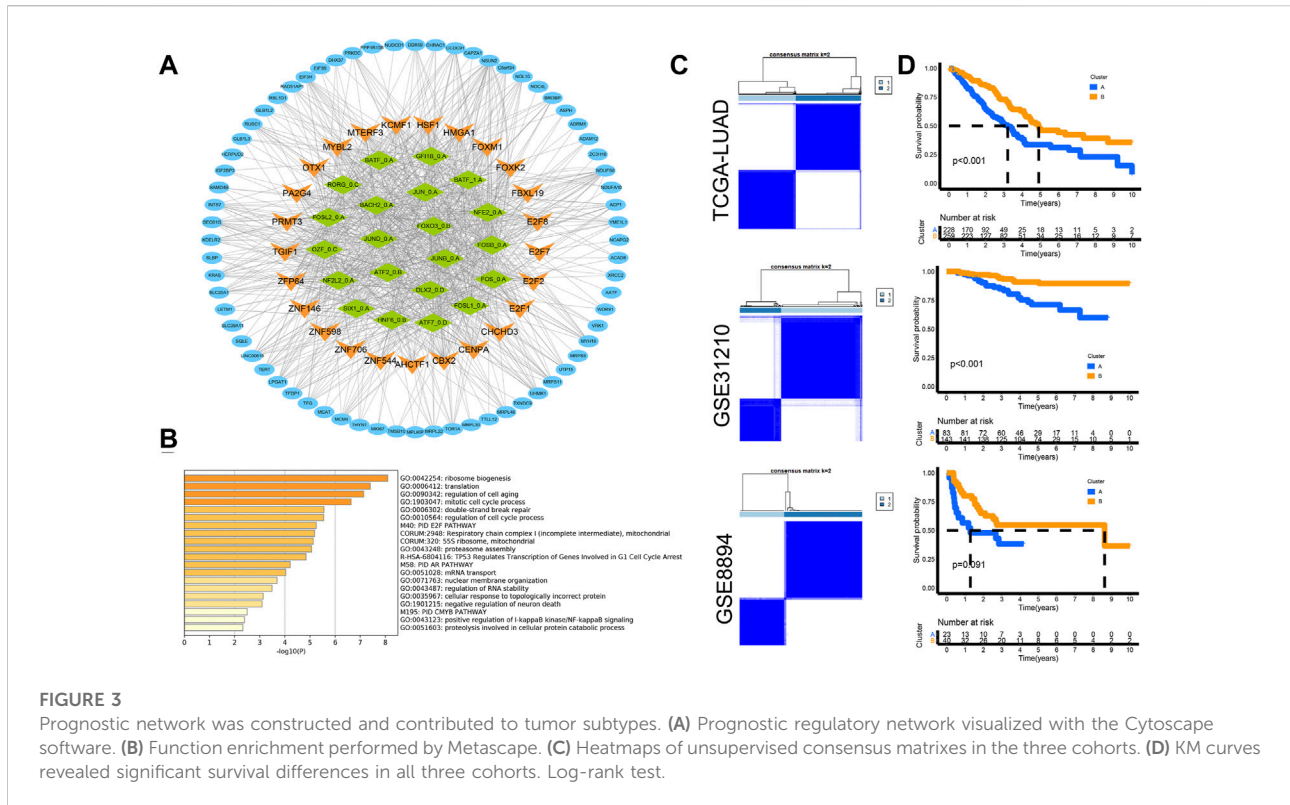
The tumor immunogenicity was analyzed in patients from high-risk and low-risk groups, which was divided by the median value of the risk score. The immunophenotype scores (IPS) of TCGA-LUAD cohort were downloaded from The Cancer Immunome Atlas database (<https://tcia.at/home>). Based on the expression status of CTLA4 and PD1 (Charoentong et al., 2017), the high-risk and low-risk groups were further classified into four subgroups: positive CTLA4 and positive PD1; positive CTLA4 and negative PD1; negative CTLA4 and positive PD1; negative CTLA4 and negative PD1. The Wilcoxon nonparametric test was used to compare the differences in IPS between high-risk and low-risk groups in each subgroup. The drug susceptibility has been explored in patients from high-risk and low-risk groups. The pRRophetic R package was applied to analyze IC50 values of six commonly used drugs (Cisplatin, Docetaxel, Erlotinib, Gefitinib, Gemcitabine, and Paclitaxel).

Integration of the enhancer-associated prognostic nomogram

The predictive efficacy of the risk score for other clinical symptoms was assessed by ROC curves for 1, 3, and 5 years of survival. The risk score and clinical features (Gender, Age, Stage, Prior-malignancy) were integrated into a nomogram using the “rms” R package. The ROC curves of the nomogram and clinical features for the 5-year survival were plotted with the survivalROC R package. The performance of the nomogram was also confirmed by both the KM and calibration curves.

Statistics

Statistical analysis was conducted with the R 4.0.3. The survival analyses were performed with the log-rank test. The comparison between the two groups was executed with the Wilcoxon test or *t*-test. *p* < 0.05 was considered to be significant.



Results

Construction of enhancer-associated regulatory network with ELMER

Both methylation and transcriptomic data of 463 primary tumor tissues and 21 normal controls from the TCGA-LUAD cohort were fed into subsequent ELMER analysis. The heatmap of methylation and transcriptomic data is plotted (Figure 2A). Distal probes are the region where enhancers are enriched. Based on the hg38 reference genome file, we first select 13,1371 distal probes with the get.feature.probe function (Supplementary Material S1). Then, the unsupervised mode is used to identify distal hypomethylated probes with the get.diff.meth function. In detail, for every distal probe, the methylation levels are sorted in all samples within the primary tumor and the normal groups separately, and those samples in the lower quintile (20% samples with the lowest methylation levels) of each group are used to identify whether the probe is hypomethylated in the tumor group, thus obtaining 10,665 distal hypomethylated probes (Supplementary Material S2) with the threshold of false discovery rate (FDR) < 0.01 and $\Delta\beta < -0.3$ (Figure 2B). Next, the GetNearGenes function is applied to identify the top ten genes closest to the upstream and downstream of the distal hypomethylated probes separately, generating probe-gene pairs (Figure 2C). Then, for each probe-gene pair, the inverse correlations between the methylation level of the probe and the

expression of the gene were tested. The top 20% and the bottom 20% of all the samples based on the probe's methylation level are extracted as the Methylated (M) group and Unmethylated (U) group. The gene expression levels between M and U groups are compared by the Mann-Whitney *U* test. 669 pairs of statistically significant probe-gene pairs with negative correlations are screened by default parameters using the get.pair function in an unsupervised mode (Supplementary Material S3). Further, the 250 bp base sequence upstream and downstream of the probes screened in the previous step are extracted, mapped to 768 motifs (Supplementary Material S4), and identified 24 significantly enriched motifs by the get.enriched.motif function (Figure 2D). Finally, based on the methylation level, the distal probes corresponding to the same motif were classified as the top 20% M group and the bottom 20% U group. A total of 127 TFs (Lambert et al., 2018) with differential expressions in the two groups and negative correlations with methylation levels are screened by using the get.TFs function in an unsupervised mode (Figure 2E), corresponding to 21 motifs (Supplementary Material S5).

Construction of prognostic network and its contribution to tumor subtypes

With the 21 motifs as links, 127 TFs and 271 target genes are screened in the enhancer-associated regulatory network.

TABLE 1 Clinical characteristics.

| Feature | TCGA-LUAD (N = 487) | GSE31210 (N = 226) | GES8894 (N = 63) |
|---------|---------------------|--------------------|------------------|
| Age | | | |
| >65 | 230 (47.23%) | 50 (22.12%) | 18 (28.57%) |
| ≤65 | 247 (50.72%) | 176 (77.88%) | 43 (68.25%) |
| Unknown | 10 (2.05%) | NA | 2 (3.17%) |
| Gender | | | |
| Male | 226 (46.41%) | 105 (46.46%) | 34 (53.97%) |
| Female | 261 (53.59%) | 121 (53.54%) | 29 (46.03%) |
| Stage | | | |
| I | 262 (53.80%) | 168 (74.34%) | NA |
| II | 114 (23.41%) | 58 (25.66%) | NA |
| III | 79 (16.22%) | NA | NA |
| IV | 25 (5.13%) | NA | NA |
| Unknown | 7 (0.01%) | NA | NA |

NA: Not Available.

According to expression data in the TCGA-LUAD cohort, 25 TFs and 80 target genes are selected as prognostic genes with the univariate analysis $p < 0.05$ (Supplementary Material S6). Then, the prognostic regulatory network is visualized with the Cytoscape software (Figure 3A). The function enrichment is performed with the Metascape (Zhou et al., 2019) webtool (<https://metascape.org/>). The results show that the regulatory network mainly affected ribosome biogenesis, translation, cell aging, cell cycle, E2F pathway, and so on (Figure 3B). Then, based on the transcriptional data of the regulatory network, unsupervised clustering analysis is conducted in the three cohorts respectively. Two stable subtypes are obtained (Figure 3C), and the KM curves reveal significant survival differences (Figure 3D).

An eight-gene enhancer-associated prognostic signature

Next, TCGA-LUAD is applied as the training set, with GSE31210 and GSE8894 as the validation sets. Relevant clinical features are presented (Table 1). With the 105 genes in the above regulatory network as the initial variables, we subsequently conduct Lasso regression analysis (Figure 4A) and Stepwise regression analysis, and finally obtain the Cox model consisting of genes. The risk score = $(0.3486 \times E2F7) + (0.3011 \times EIF3H) + (0.2536 \times INTS1) + (0.3019 \times LPGAT1) + (0.3078 \times MCAT) + (0.3006 \times SEC61G) + (-0.6044 \times SS18L1) + (-0.2837 \times TNYN1)$. The forest plot shows the hazard ratio (HR) of each gene, and the model's concordance index reached 0.7 (Figure 4B). Detailed results of variables within the model are shown in Table 2. The AUC of ROC curves for 1-, 3-, and 5-years survival are 0.764,

0.709, and 0.635 in the TCGA-LUAD cohort, and ranged from 0.633 to 0.71 in the independent validation cohorts GSE31210 and GSE8894 (Figure 4C). KM curves show significant differences between the high-risk and low-risk groups in all three cohorts (Figure 4D). The expression heatmaps of eight genes involved in the model and the risk curves of patients were also visualized. With the increase of the risk scores, the survival statuses become worse, and the expression levels of SS18L1 and THYN1 are gradually decreased, while the expression levels of the other six genes are gradually increased (Figure 4E). The results are consistent with the HR value of each gene in the model. Based on the results of multi-cox prognostic analyses, the risk score is an indicator independent of other factors ($p < 0.05$) for predicting the survival in all three cohorts (Figure 5A). Correlation analysis with clinical traits shows that with the increase of tumor stages and degrees of metastasis, the risk scores exhibit a gradually increasing trend, and the risk scores of patients with PD responses to the primary therapy exhibit significant improvement compared to those with complete response (CR), partial response (PR), and stable disease (SD) responses (Figure 5B).

Differences in TMB, TIDE, and immune micro-environment

TMB score represents the density of non-synonymous mutation distribution in the protein-coding region, which is calculated with the number of non-synonymous mutation sites/the total length of exons. TMB is an effective index to predict the response of immunotherapy, and higher TMB indicates a better immune response. TIDE is the

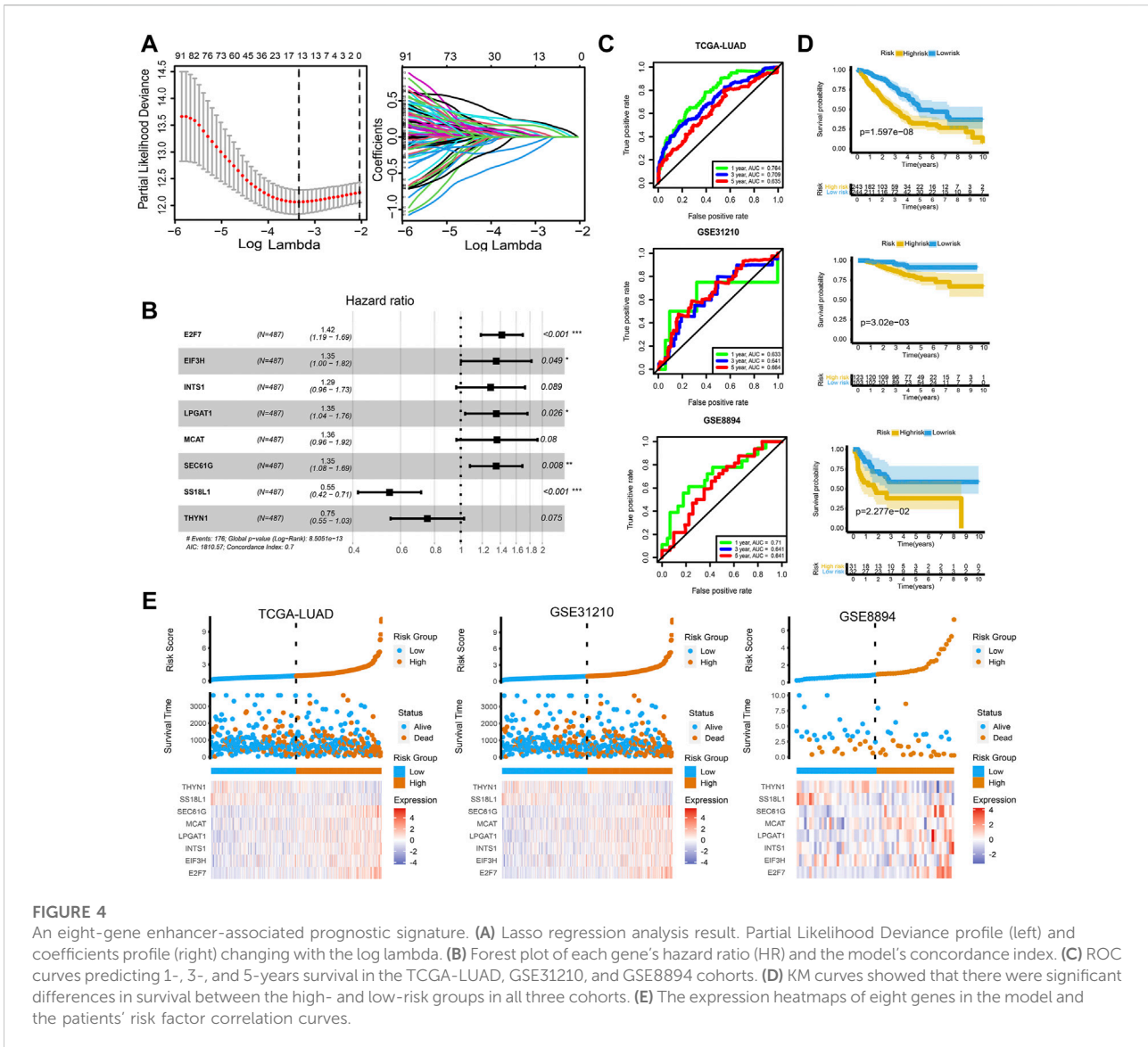
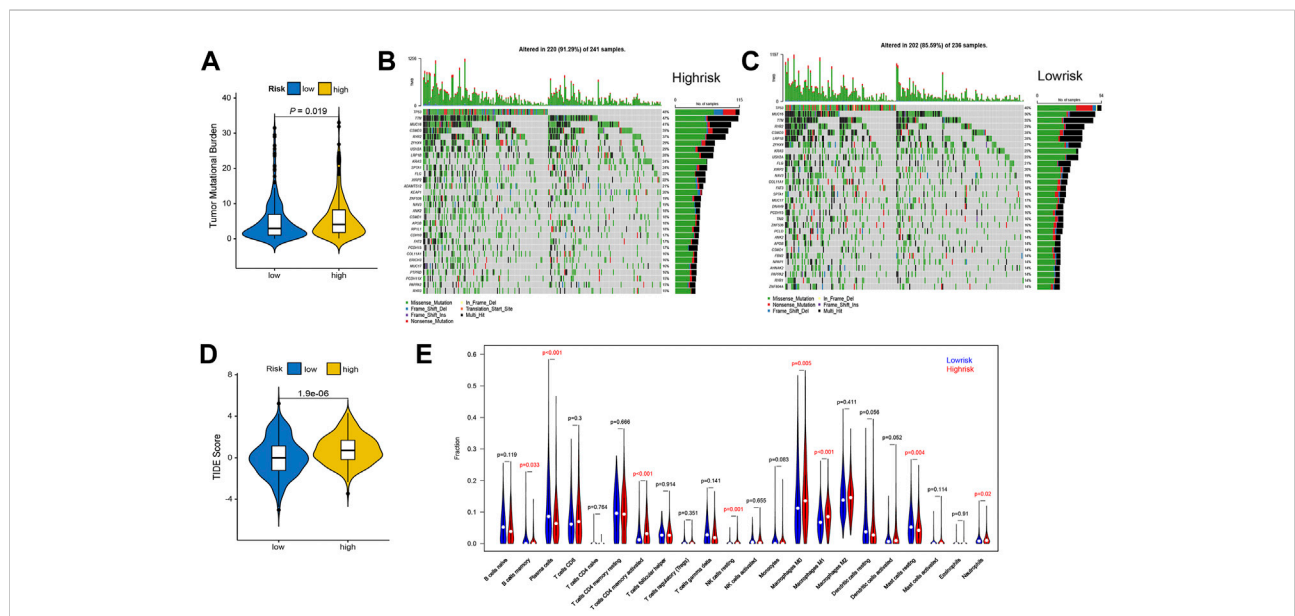
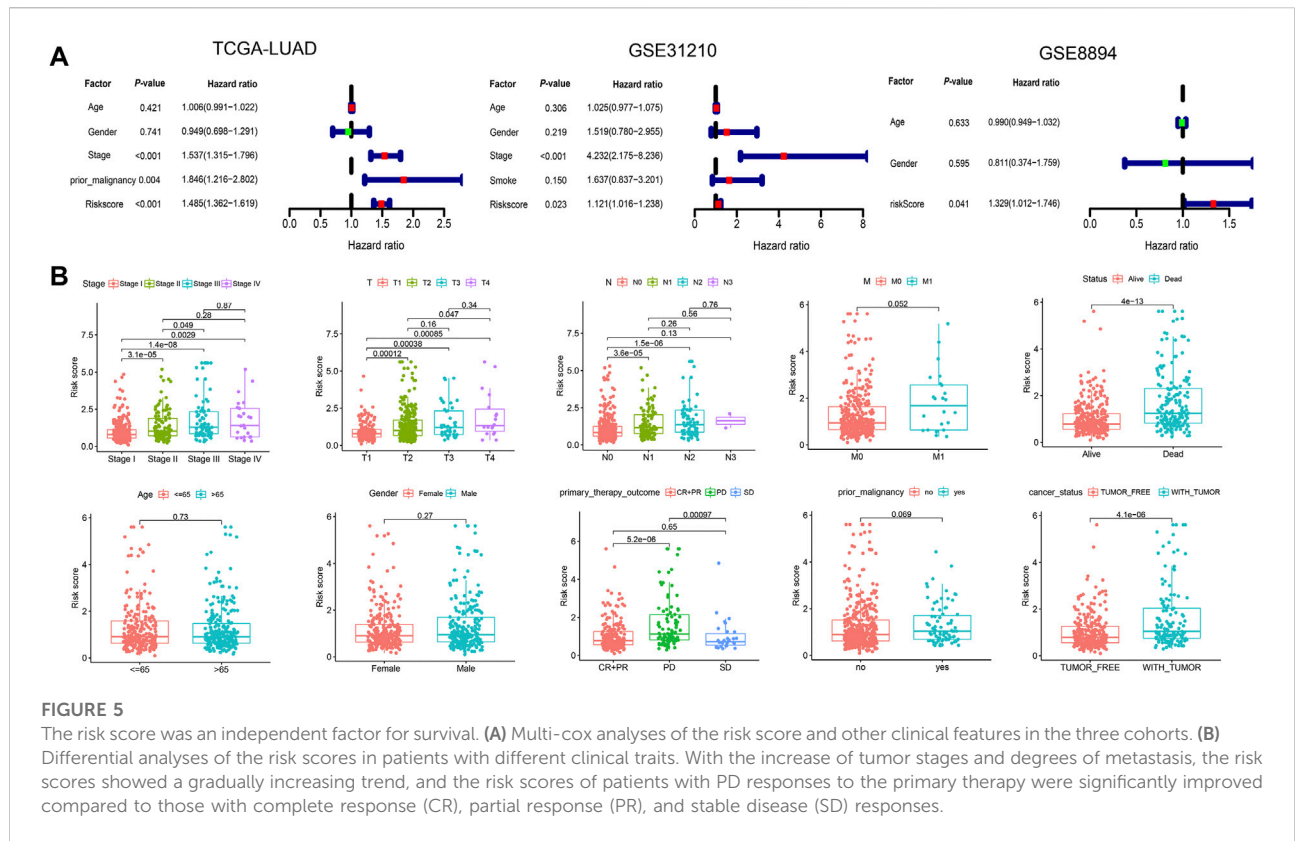


TABLE 2 Detailed results of variables within the enhancer-associated signature.

| ID | Coef | HR | HR.95 L | HR.95H | p-value |
|--------|---------|--------|---------|--------|----------|
| E2F7 | 0.3486 | 1.4171 | 1.1862 | 1.6930 | 1.22E-04 |
| EIF3H | 0.3011 | 1.3514 | 1.0013 | 1.8239 | 4.90E-02 |
| INTS1 | 0.2536 | 1.2886 | 0.9622 | 1.7258 | 8.89E-02 |
| LPGAT1 | 0.3019 | 1.3524 | 1.0376 | 1.7627 | 2.55E-02 |
| MCAT | 0.3078 | 1.3604 | 0.9639 | 1.9200 | 7.99E-02 |
| SEC61G | 0.3006 | 1.3507 | 1.0811 | 1.6876 | 8.13E-03 |
| SS18L1 | -0.6044 | 0.5464 | 0.4179 | 0.7144 | 9.90E-06 |
| THYN1 | -0.2837 | 0.7530 | 0.5511 | 1.0287 | 7.47E-02 |

computational algorithm for assessing tumor micro-environment from gene expression profiles. TIDE involves a set of gene expression markers to evaluate tumor immune evasion, including dysfunction of tumor-infiltrating cytotoxic T lymphocytes (CTL) and exclusion of CTL by immunosuppressive factors. TIDE is a quantitative index of immune escape, and higher TIDE indicates more serious the immune escape of the tumor.

To explore the characteristics of the immune microenvironment in patients with high and low risks, we compare the TMB values, TIDE scores, and 22 types of immune cell infiltration in these two groups. The high-risk group shows significantly higher TMB values (Figure 6A), and



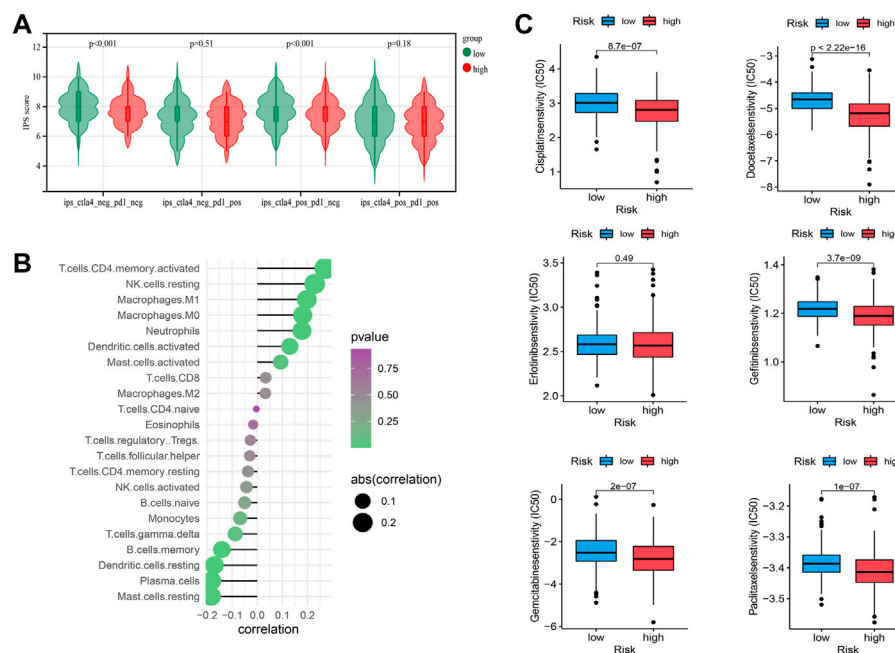


FIGURE 7

Tumor immunogenicity and drug susceptibility. **(A)** IPS score in low-risk and high-risk groups, which were classified into four subgroups: positive CTLA4 and positive PD1; positive CTLA4 and negative PD1; negative CTLA4 and positive PD1; negative CTLA4 and negative PD1, respectively. **(B)** The correlation between risk scores and various immune cell infiltration. **(C)** The IC50 values of Cisplatin, Docetaxel, Erlotinib, Gefitinib, Gemcitabine, and Paclitaxel in low-risk and high-risk groups.

higher percentages of gene mutations (Figure 6B), compared to that of the low-risk group (Figure 6C). The high-risk group also presents higher TIDE scores than the low-risk group (Figure 6D). Eight of the 22 immune cell types show statistical differences in the infiltration degree. Among them, macrophage M0, macrophage M1, and T cell CD4 memory activated significantly increase in the high-risk group, while plasma cells and mast cells resting decrease significantly (Figure 6E). T cell CD4 memory activated exhibits a significant positive correlation with the risk score, while mast cells resting presents a significant negative correlation with the risk score (Figure 7B).

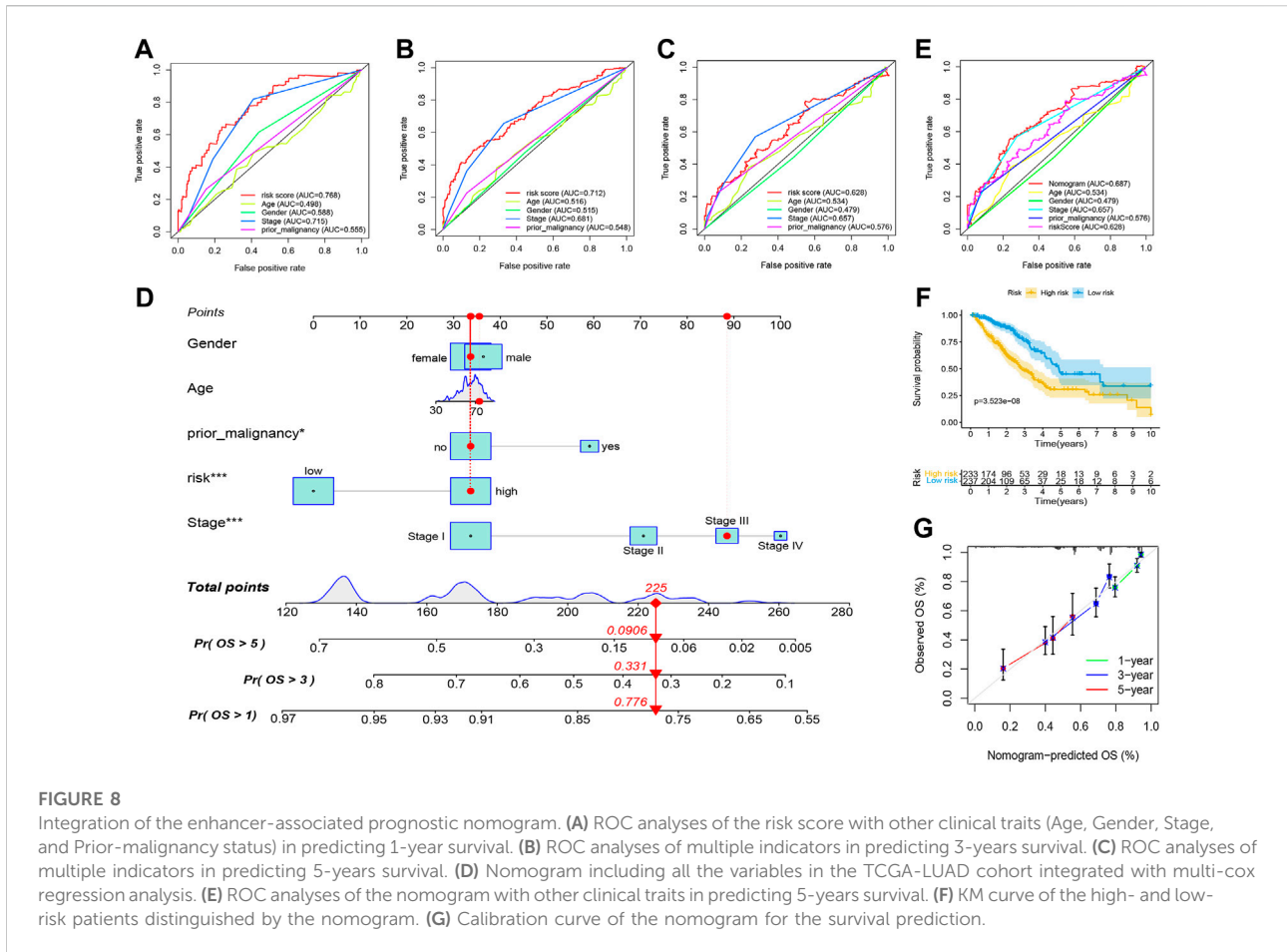
Tumor immunogenicity and drug susceptibility

The IPS is included for evaluating tumor immunogenicity. Higher IPS suggested higher immunogenicity, exhibiting a potential higher response rate to immunotherapy. The IPS score of the low-risk group is significantly higher than that of the high-risk group in subgroups of positive CTLA4 and positive PD1, negative CTLA4 and negative PD1 (Figure 7A). The results suggest that the patients in the low-risk group show a better response to immunotherapy.

The IC50 values of six commonly used drugs (Cisplatin, Docetaxel, Erlotinib, Gefitinib, Gemcitabine and Paclitaxel) are calculated in low-risk and high-risk groups. For all six drugs except for Erlotinib, LC50 is significantly higher in low-risk groups than that of high-risk groups (Figure 7C). It suggests that the patients in the high-risk group may be more sensitive to these drugs, which is accompanied by the high potential of immunosuppression.

An integrated enhancer-associated prognostic nomogram

We first compare the predictive accuracy of the risk score with various clinical traits (Age, Gender, Stage, and Prior-malignancy status). ROC analyses of multiple indicators show that the risk score is more accurate compared to other clinical traits in predicting 1- and 3-years survival (Figure 6B–Figure 8A), and slightly lower than Stage in predicting 5-years survival (Figure 8C). Then, a nomogram including all the variables in the TCGA-LUAD cohort is integrated with multi-cox regression analysis (Figure 8D). The nomogram exhibited the highest accuracy in the ROC curve for predicting 5-years survival, higher than the risk score alone (Figure 8E). And the performance is also validated with the KM curve (Figure 8F) and the calibration curve (Figure 8G).



Discussion

Lung cancer has been a common type of cancer (Mao et al., 2016). Its prognosis has been generally poor, especially for the advanced stages (Jones and Baldwin, 2018). As a huge burden on society (Bade and Dela Cruz, 2020; Wu et al., 2021), the high incidence and mortality of lung cancer put forward higher and more urgent demands for its early diagnosis and treatment (Rodriguez-Canales et al., 2016). With the development of molecular biology tools, the genomic information of lung cancer has been better profiled at the molecular level, which provided new treatment options and improved outcomes for patients (Parikh, 2019). Several studies have tried to find diagnostic and prognostic prediction markers for lung cancers. One prospective cohort study included 426 patients with complete surgical resection of stages I to III LUAD reported that a computational machine-learning prediction model integrating genomic and clinicopathologic features could better predict the risk of recurrence, compared with the current TNM system. It would provide recommendations for adjuvant therapy after surgical resection of LUAD (Jones et al., 2021). Another study performed on LUAD patients in eastern

China detected molecular alterations with a customized DNA panel. Some of the specific mutations may make effects on the efficacy of targeted therapies, CCAAT enhancer binding protein alpha (CEBPA) mutations affected the efficacy of EGFR-tyrosine kinase inhibitors. The erb-b2 receptor tyrosine kinase 2 (ERBB2), CEBPA and transcription factor 7 like 2 (TCF7L2) mutated tumors tend to have higher TMB. The targeted DNA panel may be helpful for personalized treatment decisions of LUAD patients (Liu et al., 2021).

For better understanding the molecular characteristics of lung cancers, several studies have involved various bioinformatic tools to comprehensively analyze the omics data based on microarray or sequencing analysis of patients. Based on the microarray datasets of three cohorts of lung cancer, a meta-analysis has been performed. There were 50 upregulated and 87 downregulated genes overlapped in three datasets, which were included in following analysis. With the protein-protein interaction (PPI) networks, 22 core genes were identified, which were all significantly associated with poor survival. Finally, KEGG pathway enrichment reanalysis screened five key genes, which exhibit a relationship with certain drugs. The identified key genes

can be candidate targets for both the treatment and prognosis of lung cancer (Wang et al., 2021). With the synthetic analysis of the transcriptome sequencing dataset and a non-coding RNA sequence dataset of small-cell lung cancer, the differentially expressed genes and miRNAs can be screened. After function enrichment, the molecular mechanisms were identified with the PPI network. Finally, 19 overlapping target genes and 32 corresponding regulatory miRNAs were screened. The bioinformatics analysis involving multi-omics data can assist in exploring the roles of target genes, miRNA, and TFs, which may better understand the potential molecular pathways (Mao et al., 2019).

After screening the potential candidate genes based on bioinformatics analysis, they can be further validated experimentally. A comprehensive bioinformatics analysis revealed that, the decreased expression of immunoglobulin superfamily member 10 (IGSF10) was associated with the shortened overall survival duration of patients with lung cancer. In subsequent experimental validation, IGSF10-knockout cells presented significantly increased proliferation and adhesion capability, revealed by MTT, colony formation assay, and Transwell assay, respectively. Further, Western blotting suggested that, the IGSF10-knockout can activate the integrin- β 1/FAK pathway, presented as the upregulated protein expression levels of integrin- β 1, phosphorylated (p)-FAK and p-AKT (Ling et al., 2020). Another study has analyzed the RNA sequencing data and revealed circXPO1, a novel circular RNA (circRNA) in LUAD. The circXPO1 was derived from a well-established cancer therapeutic target, XPO1, which was highly expressed in LUAD tissues compared with paired controls. High circXPO1 expression was correlated with worse overall survival. Mechanically, circXPO1 could bind with IGF2BP1 and enhance CTNNB1 mRNA stability, and subsequently promote LUAD progression (Huang et al., 2020). A similar study verified that circ-CAMK2A enhanced LUAD metastasis by regulating the miR-615-5p/fibronectin one pathway. Circ-CAMK2A upregulated the expression level of fibronectin one by sponging miR-615-5p, thus promoting MMP2 and MMP9 expression to stimulate the metastasis of LUAD (Du et al., 2019). In short, the combination of bioinformatics analysis and experimental verification can better clarify the significance of certain target biomarkers or promising pathways.

In our study, the comprehensive multi-omics analysis has also been applied to screen the target genes, which may be significant to prognosis prediction of lung cancer. The transcriptome and methylation data of the TCGA-LUAD cohort involving 463 primary tumor tissues and 21 normal controls were obtained for subsequent ELMER analysis. A total of 127 TFs corresponding to 21 motifs and 271 target genes were screened for constructing the subsequent enhancer-associated regulatory network. 25 TFs and 80 target genes were selected as prognostic genes with the univariate analysis $p < 0.05$. With TCGA-LUAD as the training set, the Cox model involving eight genes was selected with LASSO regression analysis and Stepwise regression analysis.

The risk score = $(0.3486 \times E2F7) + (0.3011 \times EIF3H) + (0.2536 \times INTS1) + (0.3019 \times LPGAT1) + (0.3078 \times MCAT) + (0.3006 \times SEC61G) + (-0.6044 \times SS18L1) + (-0.2837 \times TNYN1)$. With the increase of the risk scores, the survival statuses became worse, as well as the clinical traits including tumor stages, metastasis degree, and treatment responses. The risk score exhibited prognostic prediction accuracy with GSE31210 and GSE8894 as the validation sets.

The multi-omics study has been proved as a good tool for the epigenetic regulation of functional enhancers. In a study on HCC, methyl-binding DNA capture sequencing was firstly performed on both tumor and control tissues. The data revealed abnormal enhancer hypermethylation patterns. Then, the single-base resolution whole-genome bisulfite sequencing (WGBS) was performed to screen enhancers with differential methylation. Then, CCAAT/enhancer-binding protein-beta (C/EBP β) enhancer was selected for further function mechanism. The survival analysis indicated that hypomethylation of C/EBP β enhancer was related to the poor prognosis of patients with HCC. The mechanism has been also investigated experimentally (Xiong et al., 2019). This study has inspired several studies to perform multi-omics analysis (Cui et al., 2021). By involving methylome, transcriptome, and 3D genomic data, the researchers comprehensively analyzed enhancer methylation regulome and identified enhancer methylation-enhancer TF-target gene expression. They found that the enhancer-regulated core TFs could further shape their enhancer methylation, thus forming the enhancer methylation-driven core transcriptional regulatory circuitries, which can be served as innovative therapy targets and prognostic risk biomarkers (Pan et al., 2022). In another study integrating ChIP-seq, RNA-seq, and WGBS data, the enhancers with differential expression and differential methylation were identified, as well as the associated differentially expressed genes. A model based on six enhancer-associated genes was constructed with regression analysis, exhibiting excellent predictive accuracy (Huang et al., 2022). In addition to screening biomarkers, the combination of epigenetic and transcriptional data can also demonstrate the mechanism. For example, the aberrant methylation of promoters and enhancers could activate critical cell cycle-related pathways and inhibit several metabolic pathways, thus affecting the progression of HCC (Huang et al., 2021). In short, the integrative analysis of multi-omics data can help us find new and more effective function targets in various diseases.

One comprehensive study has summarized more than 30 bioinformatics approaches for enhancer identification. With the advances in biological technologies, several data resources have been involved for screening enhancers, such as evolutionary conservation data, histone marks, Open chromatin, Transcription factor-binding sites, Sequencing features, Screening data, and eRNA expression. These data types can be combined in different ways to generate feature vectors that describe DNA regions. After feature selection, the feature vectors feed computational models that make

decisions using unsupervised and/or supervised algorithms, such as Clustering, Classification, Graphical models, and Regression. The outcome is a list of identified enhancer regions. However, one of the major challenges is how to assess the correctness of predicted enhancers, because there is no large, sufficiently comprehensive, and experimentally validated enhancer set for humans. One possible way of validation is to link the predicted enhancers to their target genes (Kleftogiannis et al., 2016). The ELMER analysis applied in this study has involved five steps, which can construct a “TFs-motifs-genes” regulatory network. The formation of a regulatory network integrating TF, motif, and target gene further guaranteed the correctness of predicted candidates. However, one limitation of this study may be the lack of experimental evidence for validation. The prognostic prediction signature for LUAD has been only validated in patients from another two LUAD cohorts.

Conclusion

This study has comprehensively analyzed the transcriptome and methylation data of a LUAD cohort. ELMER analysis has been performed to screen motifs, motif-associated TFs, and target genes. The “TFs-motifs-genes” regulatory network was constructed. After regression analysis, the Cox model involving eight genes was constructed. The enhancer-associated prognostic gene signature can be applied as a risk score for predicting the survival status of patients. With the increase of the risk scores, both the survival statuses and clinical traits showed a worse tendency in patients with lung adenocarcinoma. The multi-omics bioinformatics analysis can be a good tool for obtaining more information at the epigenetic level.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: www.ncbi.nlm.nih.gov/geo/, GSE31210, GSE8894.

Author contributions

SH, SC, and LL performed the conception and design of this study. DZ and JG collected the datasets from databases and contributed to the organization and typesetting of the figures. SH

and SC conducted the bioinformatic analyses. SH and LL wrote the initial edition of the manuscript. SC and LL checked the statistical results. All the authors read and approved the submitted article.

Conflict of interest

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

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/fgene.2022.1008602/full#supplementary-material>

SUPPLEMENTARY MATERIAL S1

131,371 distal probes selected with the `get.feature.probe` function.

SUPPLEMENTARY MATERIAL S2

10,665 distal hypomethylated probes identified with the `get.diff.meth` function in the tumor group.

SUPPLEMENTARY MATERIAL S3

669 pairs of statistically significant probe-gene pairs with negative correlations were screened by default parameters using the `get.pair` function in an unsupervised mode.

SUPPLEMENTARY MATERIAL S4

768 motifs enriched within the 250bp base sequence upstream and downstream of the probes in the 669 probe-gene pairs.

SUPPLEMENTARY MATERIAL S5

127 TFs with differential expressions in the two groups and negative correlations with methylation levels were screened by using the `get.TFs` function.

SUPPLEMENTARY MATERIAL S6

25 TFs and 80 target genes were selected as prognostic genes with the univariate analysis p value <0.05 .

References

- Adhikary, S., Roy, S., Chacon, J., Gadad, S. S., and Das, C. (2021). Implications of enhancer transcription and eRNAs in cancer. *Cancer Res.* 81 (16), 4174–4182. doi:10.1158/0008-5472.Can-20-4010
- Bade, B. C., and Dela Cruz, C. S. (2020). Lung cancer 2020. *Clin. Chest Med.* 41 (1), 1–24. doi:10.1016/j.ccm.2019.10.001

- Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12
- Chen, H., and Liang, H. (2020). A high-resolution map of human enhancer RNA loci characterizes super-enhancer activities in cancer. *Cancer Cell* 38 (5), 701–715. e705. doi:10.1016/j.ccell.2020.08.020
- Chen, J., Yang, H., Teo, A. S. M., Amer, L. B., Sherbaf, F. G., Tan, C. Q., et al. (2020). Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* 52 (2), 177–186. doi:10.1038/s41588-019-0569-6
- Cui, D., Li, W., Jiang, D., Wu, J., Xie, J., and Wu, Y. (2021). Advances in multi-omics applications in HBV-associated hepatocellular carcinoma. *Front. Med.* 8, 754709. doi:10.3389/fmed.2021.754709
- Denisenko, T. V., Budkevich, I. N., and Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma. *Cell Death Dis.* 9 (2), 117. doi:10.1038/s41419-017-0063-y
- Du, J., Zhang, G., Qiu, H., Yu, H., and Yuan, W. (2019). The novel circular RNA circ-CAMK2A enhances lung adenocarcinoma metastasis by regulating the miR-615-5p/fibronectin 1 pathway. *Cell Mol. Biol. Lett.* 24, 72. doi:10.1186/s11658-019-0198-1
- Field, A., and Adelman, K. (2020). Evaluating enhancer function and transcription. *Annu. Rev. Biochem.* 89, 213–234. doi:10.1146/annurev-biochem-011420-095916
- Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., et al. (2017). GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database* 2017. doi:10.1093/database/bax028
- Helgason, P. L., Parolia, A., Liu, H. H., and Helgason, C. D. (2016). DNA methylation at enhancer regions: Novel avenues for epigenetic biomarker development. *Front. Biosci.* 21 (2), 430–446. doi:10.2741/4399
- Herz, H. M. (2016). Enhancer deregulation in cancer and other diseases. *Bioessays* 38 (10), 1003–1015. doi:10.1002/bies.201600106
- Huang, P., Xu, M., Han, H., Zhao, X., Li, M. D., and Yang, Z. (2021). Integrative analysis of epigenome and transcriptome data reveals aberrantly methylated promoters and enhancers in hepatocellular carcinoma. *Front. Oncol.* 11, 769390. doi:10.3389/fonc.2021.769390
- Huang, P., Zhang, B., Zhao, J., and Li, M. D. (2022). Integrating the epigenome and transcriptome of hepatocellular carcinoma to identify systematic enhancer aberrations and establish an aberrant enhancer-related prognostic signature. *Front. Cell Dev. Biol.* 10, 827657. doi:10.3389/fcell.2022.827657
- Huang, Q., Guo, H., Wang, S., Ma, Y., Chen, H., Li, H., et al. (2020). A novel circular RNA, circXPO1, promotes lung adenocarcinoma progression by interacting with IGF2BP1. *Cell Death Dis.* 11 (12), 1031. doi:10.1038/s41419-020-03237-8
- Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1
- Jones, G. D., Brandt, W. S., Shen, R., Sanchez-Vega, F., Tan, K. S., Martin, A., et al. (2021). A genomic-pathologic annotated risk model to predict recurrence in early-stage lung adenocarcinoma. *JAMA Surg.* 156 (2), e205601. doi:10.1001/jamasurg.2020.5601
- Jones, G. S., and Baldwin, D. R. (2018). Recent advances in the management of lung cancer. *Clin. Med.* 18 (2), s41–s46. doi:10.7861/clinmedicine.18-2-s41
- Kleftogiannis, D., Kalnis, P., and Bajic, V. B. (2016). Progress and challenges in bioinformatics approaches for enhancer identification. *Brief. Bioinform.* 17 (6), 967–979. doi:10.1093/bib/bbv101
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22 (3), 568–576. doi:10.1101/gr.129684.111
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. *Cell* 172 (4), 650–665. doi:10.1016/j.cell.2018.01.029
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034
- Li, Y., Gu, J., Xu, F., Zhu, Q., Chen, Y., Ge, D., et al. (2021). Molecular characterization, biological function, tumor microenvironment association and clinical significance of m6A regulators in lung adenocarcinoma. *Brief. Bioinform.* 22 (4). doi:10.1093/bib/bbaa225
- Ling, B., Liao, X., Huang, Y., Liang, L., Jiang, Y., Pang, Y., et al. (2020). Identification of prognostic markers of lung cancer through bioinformatics analysis and *in vitro* experiments. *Int. J. Oncol.* 56 (1), 193–205. doi:10.3892/ijo.2019.4926
- Liu, J., Xu, W. Y., Ye, M., Liu, Z., and Li, C. (2021). Genetic alteration profiling of Chinese lung adenocarcinoma and its effect on targeted therapy efficacy. *Front. Oncol.* 11, 726547. doi:10.3389/fonc.2021.726547
- Mao, Y., Xue, P., Li, L., Xu, P., Cai, Y., Chu, X., et al. (2019). Bioinformatics analysis of mRNA and miRNA microarray to identify the key miRNA-gene pairs in small-cell lung cancer. *Mol. Med. Rep.* 20 (3), 2199–2208. doi:10.3892/mmr.2019.10441
- Mao, Y., Yang, D., He, J., and Krasna, M. J. (2016). Epidemiology of lung cancer. *Surg. Oncol. Clin. N. Am.* 25 (3), 439–445. doi:10.1016/j.soc.2016.02.001
- Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118
- Mazieres, J., Drilon, A., Lusque, A., Mhanna, L., Cortot, A. B., Mezquita, L., et al. (2019). Immune checkpoint inhibitors for patients with advanced lung cancer and oncogenic driver alterations: Results from the IMMUNOTARGET registry. *Ann. Oncol.* 30 (8), 1321–1328. doi:10.1093/annonc/mdz167
- Pan, X., Li, X., Sun, J., Xiong, Z., Hu, H., Ning, S., et al. (2022). Enhancer methylation dynamics drive core transcriptional regulatory circuitry in pan-cancer. *Oncogene* 41 (26), 3474–3484. doi:10.1038/s41388-022-02359-x
- Parikh, A. R. (2019). Lung cancer genomics. *ama* 48 (1), 78–83. doi:10.5644/ama2006-124.244
- Rodriguez-Canales, J., Parra-Cuentas, E., and Wistuba, II (2016). Diagnosis and molecular classification of lung cancer. *Cancer Treat. Res.* 170, 25–46. doi:10.1007/978-3-319-40389-2_2
- Ruiz-Cordero, R., and Devine, W. P. (2020). Targeted therapy and checkpoint immunotherapy in lung cancer. *Surg. Pathol. Clin.* 13 (1), 17–33. doi:10.1016/j.path.2019.11.002
- Silva, T. C., Coetzee, S. G., Gull, N., Yao, L., Hazelett, D. J., Noushmehr, H., et al. (2019). ELMER v.2: An R/bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics* 35 (11), 1974–1977. doi:10.1093/bioinformatics/bty902
- Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. *Nat. Rev. Cancer* 16 (8), 483–493. doi:10.1038/nrc.2016.62
- Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A., et al. (2017). ChAMP: Updated methylation analysis pipeline for illumina BeadChips. *Bioinformatics* 33 (24), 3982–3984. doi:10.1093/bioinformatics/btx513
- Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3
- Wang, Y., Zhou, Z., Chen, L., Li, Y., Zhou, Z., and Chu, X. (2021). Identification of key genes and biological pathways in lung adenocarcinoma via bioinformatics analysis. *Mol. Cell Biochem.* 476 (2), 931–939. doi:10.1007/s11010-020-03959-5
- Wang, Z., Zhang, Q., Zhang, W., Lin, J. R., Cai, Y., Mitra, J., et al. (2018). Hedd: Human enhancer disease database. *Nucleic Acids Res.* 46 (D1), D113–d120. doi:10.1093/nar/gkx988
- Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170
- Wu, F., Wang, L., and Zhou, C. (2021). Lung cancer in China: Current and prospect. *Curr. Opin. Oncol.* 33 (1), 40–46. doi:10.1097/cco.0000000000000703
- Xiong, L., Wu, F., Wu, Q., Xu, L., Cheung, O. K., Kang, W., et al. (2019). Aberrant enhancer hypomethylation contributes to hepatic carcinogenesis through global transcriptional reprogramming. *Nat. Commun.* 10 (1), 335. doi:10.1038/s41467-018-08245-z
- Yi, M., Li, A., Zhou, L., Chu, Q., Luo, S., and Wu, K. (2021). Immune signature-based risk stratification and prediction of immune checkpoint inhibitor's efficacy for lung adenocarcinoma. *Cancer Immunol. Immunother.* 70 (6), 1705–1719. doi:10.1007/s00262-020-02817-z
- Zhang, C., Zhang, Z., Sun, N., Zhang, Z., Zhang, G., Wang, F., et al. (2020). Identification of a costimulatory molecule-based signature for predicting prognosis risk and immunotherapy response in patients with lung adenocarcinoma. *Oncimmunology* 9 (1), 1824641. doi:10.1080/2162402x.2020.1824641
- Zhang, G., Shi, J., Zhu, S., Lan, Y., Xu, L., Yuan, H., et al. (2018). DiseaseEnhancer: A resource of human disease-associated enhancer catalog. *Nucleic Acids Res.* 46 (D1), D78–d84. doi:10.1093/nar/gkx920
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6