76820_Auto_EditedC.docx

Retrospective Study

Development and validation of an epithelial–mesenchymal transition-related gene signature for predicting prognosis

Zhou DH *et al*. Based on 884 cases of LUAD

De-Hua Zhou, Qian-Cheng Du, Zheng Fu, Xin-Yu Wang, Ling Zhou, Jian Wang, Cheng-Kai Hu, Shun Liu, Jun-Min Li, Meng-Li Ma, Hua Yu

Abstract

BACKGROUND

Currently, there are many therapeutic methods for lung adenocarcinoma (LUAD), but the 5-year survival rate is still only 15% at later stages. Epithelial–mesenchymal transition (EMT) has been shown to be closely associated with local dissemination and subsequent metastasis of solid tumors. However, the role of EMT in the occurrence and development of LUAD remains unclear.

AIM

To further elucidate the value of EMT-related genes in LUAD prognosis.

METHODS

Univariate, least absolute shrinkage and selection operator, and multivariate Cox regression analyses were applied to establish and validate a new EMT-related gene signature for predicting LUAD prognosis. The risk model was evaluated by Kaplan–Meier survival analysis, principal component analysis, and functional enrichment

analysis and was used for nomogram construction. The potential structures of drugs to which LUAD is sensitive were discussed with respect to EMT-related genes in this model.

RESULTS

Thirty-three differentially expressed genes related to EMT were found to be highly associated with overall survival (OS) by using univariate Cox regression analysis (log2FC ≥ 1, false discovery rate < 0.001). A prognostic signature of 7 EMT-associated genes was developed to divide patients into two risk groups by high or low risk scores. Kaplan–Meier survival analysis showed that the OS of patients in the high-risk group was significantly poorer than that of patients in the low-risk group (*P* < 0.05). Multivariate Cox regression analysis showed that the risk score was an independent risk factor for OS (HR > 1, *P* < 0.05). The results of receiver operator characteristic curve analysis suggested that the 7-gene signature had a perfect ability to predict prognosis (all area under the curves > 0.5).

CONCLUSION

The EMT-associated gene signature classifier could be used as a feasible indicator for predicting OS.

Key Words: Lung adenocarcinoma; Epithelial–mesenchymal transition; Gene signature; Overall survival

Zhou DH, Du QC, Fu Z, Wang XY, Zhou L, Wang J, Hu CK, Liu S, Li JM, Ma ML, Yu H. Development and validation of an epithelial–mesenchymal transition-related gene signature for predicting prognosis. *World J Clin Cases* 2022; In press

Core Tip: Lung cancer is one of the major causes of death associated with malignancy, and lung cancer is associated with approximately 2 million new cases and 1.76 million

deaths every year. Although some reports have shown that suppression or knockdown of some genes in lung adenocarcinoma (LUAD) could reverse the epithelial–mesenchymal transition (EMT) process that inhibits tumor occurrence and metastasis, the correlations of these genes with overall survival in patients with LUAD remain largely unclear. A cohort of 884 LUAD patients was used to construct and validate a novel predictive model comprising a 7-EMT-related gene signature, which can be used to predict the prognosis of LUAD.

INTRODUCTION

Worldwide, lung cancer is one of the major causes of death associated with malignancy, and lung cancer is associated with approximately 2 million new cases and 1.76 million deaths every year[1]. Approximately 80%-85% of lung carcinomas are non-small cell lung cancer, and the two major histopathological types are lung adenocarcinoma (LUAD) and lung squamous cell carcinoma[2]. Among patients with lung cancer, 40% have LUAD, which has a 5-year survival rate of only 15% at later stages[3]. Currently, there are many therapeutic methods for LUAD, including curative resection, radiotherapy, chemotherapy, targeted therapy, and immunotherapy. Although the therapeutic prognosis of lung cancer has improved significantly because of advances in technology and selection strategies, the prognosis for lung cancer patients remains poor, as 70%-80% of patients are diagnosed in advanced stages[4,5]. In modern clinical oncology, histopathology is commonly effective in predicting the prognosis of lung cancer patients; however, its value is limited, as individual distinctions in patients with the identical pathological characteristics result in different conclusions. Therefore, more effective biomarkers for the early diagnosis of LUAD and more accurate prognostic predictions are necessary.

Epithelial–mesenchymal transition (EMT) is a multistep process in which morphological changes occur and is involved in neoplastic invasion, metastasis, embryogenesis and wound healing; during EMT, epithelial cells acquire mesenchymal characteristics and gradually lose their epithelial features[6]. EMT has been shown to be closely associated with local dissemination and subsequent metastasis of solid tumors[7]. Exosomes also frequently facilitate communication between cells and the extracellular matrix, complete the EMT process, and increase the transformation of nontumor stem cells into tumor stem cells, which accelerates the development of cancer[8]. There are many EMT states that can occur in cancer, and these states contribute differently to carcinogenesis, invasion, and metastasis. Mixed EMT states have a higher propensity for metastasis[9]. Pinin was shown by Qiao et al[10] to induce EMT by controlling m6A modification, and it might also be a potential anticancer target for hepatocellular

carcinoma treatment. Previous studies suggested that highly proliferative non-EMT cells are sensitive to chemotherapy, and recurrent EMT-derived metastases were observed after treatment[11]. However, the underlying mechanism of EMT in the occurrence and development of non-small cell lung cancer, especially LUAD, remains largely unclear. Therefore, understanding the roles of EMT-related genes in LUAD diagnosis and prognostic prediction might be useful for identifying prognostic targets.

However, studies based on the transcriptome that systematically assess the EMT-related gene signature and its correlation with overall survival (OS) in LUAD patients remains rare. In this study, we explored available gene expression data from LUAD patient samples (n = 497) and extracted a total of 54774 candidate genes. In addition, GSEA identified 612 EMT-related functional gene sets that presented differential expression between tumor samples and normal samples. Then, we became the first to develop a novel prognostic model composed of differentially expressed EMT-related genes based on LUAD samples. This model based on external data (n = 443) was further validated to be an independent prognostic factor related to OS in LUAD patients.

MATERIALS AND METHODS

Baseline characteristics

The LUAD dataset, including whole-transcriptome RNA sequencing data and the corresponding clinical data of LUAD patients up to July 15, 2021, was downloaded from The Cancer Genome Atlas (https://cancergenome.nih.gov/). The clinical data of the LUAD patients included sex, grade, age, stage, and tumor, node and metastasis (TNM) classification. Data normalization was carried out utilizing the limma (Linear Models for Microarray Data) package in the R environment (version 4.1.0,2021–05-18, R Foundation for statistical computing, Vienna, Austria). Gene expression profiles were obtained using raw data, and log2 normalization was performed for each candidate gene for further statistical analysis. If one gene occupied more than one row, we took the mean of its expression value. All genes with an expression value of 0 were deleted from the sample. RNA-seq data and the corresponding clinical baseline characteristics

for 443 LUAD patient samples were collected from a Gene Expression Omnibus dataset (GSE68465, http://www.ncbi.nlm.nih.gov/geo/). All 443 candidate patients underwent surgery. Normalized count values were utilized. All the data in this article obtained from The Cancer Genome Atlas and the Gene Expression Omnibus dataset were open access and freely available. Therefore, the study did not need approval from the local ethics committee. All methods were carried out following the related policies, regulations and guidelines provided by The Cancer Genome Atlas, Gene Expression Omnibus and other databases for analysis of their data and presentation of the related findings.

The hallmark gene sets, including 612 specific gene sets, comprised significantly expressed signature genes obtained from Molecular Signatures Database (version 7.4) gene sets to show perfectly defined biological states or processes. Gene set enrichment analysis (GSEA) was carried out using GSEA software (v4.1.0) to establish whether a fixed set of genes exhibited a statistically significant difference between certain phenotypes[12]. We acquired nine EMT-related gene sets: SARRIO_EPITHELIAL_MESENCHYMAL_TRANSITION_DN, GOBP_CARDIAC_EPITHELIAL_TO_MESENCHYMAL_TRANSITION, GOBP_EPITHELIAL_TO_MESENCHYMAL_TRANSITION, GOBP_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_INVOLVED_IN_ENDOCARDIAL_CUSHION_FORMATION, GOBP_POSITIVE_REGULATION_OF_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_INVOLVED_IN_ENDOCARDIAL_CUSHION_FORMATION, GOBP_REGULATION_OF_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_INVOLVED_IN_ENDOCARDIAL_CUSHION_FORMATION, HOLLERN_EMT_BREAST_TUMOR_DN, JECHLINGER_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_DN AND SARRIO_EPITHELIAL_MESENCHYMAL_TRANSITION_UP, that were obviously enriched with a normalized P value of < 0.05 (Table 1, Figure 1). Then, 612 EMT-associated genes were identified by filtering according to GSEA.

Construction and validation of a new EMT-associated gene signature

To identify differentially expressed genes (DEGs) between cancer tissues and paracancerous tissues, we utilized the "limma" package in the R environment. The criteria used to identify DEGs in The Cancer Genome Atlas-LUAD cohort were log2FC ≥ 1 and false discovery rate (FDR) < 0.001. Univariate Cox regression analysis was carried out to evaluate OS-related genes (*P* values < 0.05). To predict correlations among DEGs, we constructed a PPI network using the online STRING database (version 11.0, https://string-db.org/). To reduce the risk of overfitting, we utilized the least absolute shrinkage and selection operator (LASSO) Cox regression model ("glmnet" R package), and the minimum criteria were selected[13,14]. The independent variable in the LASSO regression analysis was the standardized expression matrix of candidate prognostic DEGs associated with EMT, and the dependent variable was the OS (or other survival) status of LUAD patients in The Cancer Genome Atlas cohort. The optimal penalty parameter lambda (λ) and the matching coefficients were confirmed by 10-fold cross-validation according to the minimum criteria in the LASSO Cox regression algorithm. Lambda (λ) was the minimum partial likelihood deviance that was used for feature selection. The linear combination of gene expression values weighted by the corresponding regression coefficients was used to calculate the risk scores of LUAD patients according to the following formula: Risk score = $e^{sum\ (expression\ value\ of\ each\ gene\ \times\ corresponding\ regression\ coefficient)}$. LUAD patients were classified into a low-risk group and a high-risk group according to the median risk score. Principal component analysis (PCA) was carried out with the "prcomp" function in the R package "stats". The t-distributed stochastic neighbor embedding (t-SNE) method in the R package "Rtsne" was performed for cluster visualization. For survival analysis based on each gene, the "surv_cutpoint" function in the "survminer" package was used to identify the best cutoff value of the risk score. The R package "survivalROC" was used to assess the predictive prognostic performance of the EMT-associated gene signature *via* time-dependent receiver operator characteristic curve (ROC) curve estimation.

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses based on the prognostic DEGs (P value < 0.05, FDR < 0.05) between the high-risk group and the low-risk group were carried out using the R package "clusterProfiler". Our GO functional enrichment analysis covered three ontologies: biological process, cellular component and molecular function. KEGG pathway enrichment analysis based on a hypergeometric algorithm was carried out to determine the significant KEGG pathways, and the results were visualized with the "ggplot2" package in the R environment. Adjusted P values were calculated by the Benjamini–Hochberg method.

Construction of a predictive nomogram

R software was used to develop a novel nomogram model containing predictors (age, risk score, and TNM stage) related to OS at 1, 2 and 3 years. Calibration curves were generated to estimate the agreement of the model between the actual outcomes and predicted outcomes for one-year OS, two-year OS and three-year OS.

Identifying potential therapeutic targets

We uploaded the prognostic DEGs with a P value of < 0.05 to CMap (https://portals.broadinstitute.org/cmap), which is a biological application database of small molecule drugs, gene expression profiles, and diseases. The CMap database was built based on DEGs in human cells treated with small molecule drugs. The enrichment score based on similarity was ultimately calculated. A positive correlation score showed that a small molecule drug increased the risk of death in OS patients. In contrast, a negative correlation score showed that a drug reduced the risk of death. Drugs with negative connectivity scores had a potential therapeutic application. We obtained the three-dimensional structures of the candidate small molecule therapeutic drugs from PubChem (https://pubchem.ncbi.nlm.nih.gov/).

Statistical analysis

Student's *t* test was performed to identify differences in gene expression between cancer tissues and adjacent paracancerous tissues. Differences between proportions were evaluated by the χ^2 test. Overall survival was evaluated by Kaplan–Meier survival analysis, and the survival of the groups was compared using the log-rank test. Univariate and multivariate Cox regression analyses were carried out to explore independent prognostic factors of OS in LUAD patients. All statistical analyses were performed using SPSS software (version: 23.0, IBM Corporation, 2015, United States) and R software. All *P* values were two-tailed, and statistical significance was defined as $P < 0.05$.

RESULTS

The overall workflow of our research in light of model development and subsequent analyses is visualized in Figure 2. All 445 patients with LUAD from The Cancer Genome Atlas cohort and 439 patients from the gene set enrichment analysis cohort (GSE68465) were finally combined. The specific clinical baseline data of these participants are presented in Table 2.

Screening for prognostic EMT-associated DEGs

EMT-associated genes (42/612, 6.86%) were differentially expressed between cancer tissues and adjacent paracancerous tissues, and univariate Cox regression analysis showed that there were 33 genes associated with OS in LUAD (Figure 3A). The heatmap showed that *ANXA8L1* and *INMT* were highly expressed in adjacent paracancerous tissues (Figure 3B). However, the forest plots showed that the expression of *ANXA8L1* was upregulated in cancer tissues, and the results of univariate Cox regression analysis showed that there was a statistically significant association between *ANXA8L1* gene expression and OS (Figure 3C). Therefore, *ANXA8L1* was excluded from further study. In addition, 7 genes (*UHRF1*, *INMT*, *ANLN*, *SKA3*, *SAPCD2*, *GJB2* and *CDCA5*) were not found in the gene expression omnibus cohort. Finally, a total of 25 EMT-related

DEGs associated with prognosis were used in subsequent analyses (all *P* values < 0.05, Figure 3B-E). The protein–protein interaction network of these 25 genes showed that *PLK1*, *NDC80* and *BUB1B* were considered the hub genes (Figure 3D). The mutual connections among the 25 genes are shown in Figure 3E.

Development of a novel model for predicting prognosis

To assess the risk probability for each patient, LASSO Cox regression analysis was performed to develop a predictive model based on prognostic genes utilizing the expression profiles of the 25 EMT-related genes indicated above. According to the optimum value of lambda (λ), 7 possible factors (*HMMR*, *PLK1*, *ASPM*, *EXO1*, *KRT16*, *SPRR1B* and *EFNA3*) were finally confirmed. Survival analysis according to the optimum threshold expression value of each gene showed that there was a significant correlation between high expression of these genes and poor prognosis (*P* value < 0.05). The risk score formula was as follows: Risk score = $e^{(0.229 \times \text{expression level of HMMR} + 0.484 \times \text{expression level of PLK1} + 0.018 \times \text{expression level of ASPM} + 0.085 \times \text{expression level of EXO1} + 0.075 \times \text{expression level of KRT16} + 0.178 \times \text{expression level of SPRR1B} + 0.086 \times \text{expression level of EFNA3})}$. The LUAD patients were classified into the high-risk group (*n* = 222) or the low-risk group (*n* = 223) on the basis of the median cutoff value (Figure 4A). Patients in the high-risk group were significantly more likely to have a more advanced TNM stage (Table 3). PCA and t-SNE analyses were carried out to ascertain the difference in the distribution between the two risk groups (Figure 4B and C). The results showed that patients with higher risk scores had shorter survival times than those with lower risk scores (Figure 4D). Kaplan–Meier analysis showed that the OS of patients in the high-risk group was significantly poorer than that of patients in the low-risk group (*P* = 1.335 × 10⁻⁶, Figure 4E). Time-dependent ROC curve analysis showed that the model for predicting prognosis had favorable performance in predicting OS [area under the curve (AUC) for one-year OS = 0.638, AUC for two-year OS = 0.653, AUC for three-year OS = 0.688; Figure 4F].

Validation of the predictive model in an external cohort

Survival analysis based on the 7 EMT-associated genes in the signature proved that these genes were associated with poorer OS. To confirm the robustness of the predictive model, the LUAD patients in the external cohort were also classified into the high-risk group and the low-risk group according to the median risk score calculated by the same formula used in the derivation cohort (Figure 5A). Patients in the high-risk group in this validation cohort were also more likely to have a more advanced TNM stage, and the evidence for this association was entirely consistent (Table 3). Moreover, the results of PCA and t-SNE analysis showed that the LUAD patients in the different risk groups were distributed in two directions (Figures 5B and 5C). Similarly, we found the number of nonsurviving patients was much higher in the high-risk group, revealing that patients with high risk scores might have poor survival outcomes (P = 3.656 × 10^{-3}, Figure 5D and E). Additionally, the AUCs for survival in the validation model were 0.678 at one year, 0.691 at two years, and 0.651 at three years (Figure 5F).

Assessment of the independent prognostic value of the 7-gene signature

The OS-associated indicators confirmed by univariate Cox regression analysis were subsequently determined utilizing multivariate Cox regression analysis. The results of univariate Cox regression analysis showed that the risk score was considerably associated with OS in both the modeling cohort and the validation cohort (modeling cohort: HR = 3.849, 95%CI: 2.421-6.120, P < 0.001; validation cohort: HR = 1.441, 95%CI: 1.198-1.735, P < 0.001; Figure 6A and C). After controlling for confounding effects, multivariate Cox regression analysis showed that the risk score was an independent predictor for OS in LUAD patients (derivation cohort: HR = 3.136, 95%CI: 1.976-4.975, P < 0.001; validation cohort: HR = 1.330, 95%CI: 1.099-1.609, P = 0.003, respectively) (Figure 6B and 6D).

Functional analyses in the two cohorts

To further elucidate the biological functions and relevant pathways associated with the risk score, GO functional enrichment analysis and KEGG pathway enrichment analysis

were carried out using the DEGs between the high-risk group and the low-risk group. The GO molecular function terms enriched with the DEGs in the modeling cohort and validation cohort were mainly associated with the cell cycle and cell division and included nuclear division, organelle fission, and mitotic nuclear division ($P < 0.05$, Figure 7A and C). Surprisingly, the results of GO functional enrichment analysis were completely consistent between the modeling cohort and validation cohort. The results of KEGG pathway enrichment analysis showed that the cell cycle pathway was enriched in the two cohorts ($P < 0.05$, Figure 7B and D).

Development and evaluation of the prognostic nomogram

A novel prognostic nomogram including the risk score and clinical risk factors was developed to evaluate one-year, two-year, and three-year OS. Combined with clinical risk factors, the risk level in the model for predicting prognosis had significant predictive ability in the nomogram (Figure 8A). The calibration plots for the novel nomogram for predicting one-year, two-year and three-year OS are shown, and all exhibited satisfactory accuracy (Figure 8B-D).

Identification of promising therapeutic targets

We uploaded 12012 DEGs with P value < 0.05, namely, 1186 downregulated genes and 4982 upregulated genes, into the CMap network tool. According to the screening criteria of an enrichment score < 0 and a P value < 0.05, 66 small molecule drugs were found to have significant antitumor effects. Among these highly significantly correlated molecules, GW-8510, menadione, 6-thioguanosine, phenoxybenzamine, medrysone, chlorpromazine, 8-azaguanine, meticrane, morantel and resveratrol were among the 10 drugs with the highest negative correlations with high-risk patients (Figure 9). The three-dimensional structure of 0175029-0000 was not found in the CMap network tool; thus, this drug was not included.

DISCUSSION

LUAD is recognized as the most common and aggressive malignant tumor with variable prognosis and is a threat to health and life[15]. Epithelial cells can lose their epithelial phenotypes, including cell polarity and connection to the basement membrane, and acquire mesenchymal phenotypes with potent migration, invasion, anti-apoptotic and extracellular matrix degradation abilities, thus acquiring important biological behaviors of migration and invasion[16]. To the best of our knowledge, no prior research has explored the correlation between EMT and LUAD growth. Surprisingly, EMT-related genes were differentially expressed in tumor and normal tissues, and these genes were substantially correlated with OS (33/42, 78.57%). This suggests that EMT might have a role in LUAD prognosis and that this risk score is useful for predicting LUAD survival. Previous studies have also reported that the malignant progression of non-small cell lung cancer is determined by EMT and PD-L1 and is related to poor survival in patients with cancer[17].

GSEA is a gene set enrichment analysis approach that combines information from several sources and levels. Using expression data for 54774 genes from 551 individuals with lung cancer, we attempted to conduct GSEA in this study and discovered significant differences in 9 functions. EMT, with a statistically significant P value, was chosen for further study based on the normalized enrichment score, size, and P value. *Via* GSEA, we concentrated on a particular function to select genes for predicting patient survival and thoroughly investigated these genes. A combination of 7 genes with prognostic significance for patients with LUAD, rather than a single gene, was found through univariate and multivariate Cox regression analyses. This selected risk signature might be slightly more focused and have greater predictive power than certain other known prognostic biomarkers, supporting clinical outcomes and serving as a useful categorization tool for LUAD patients. Using the TCGA LUAD dataset, genes associated with EMT were identified, and their expression was compared to that in adjacent noncancerous tissues. High risk scores were linked to worse survival according to Kaplan–Meier analysis. These findings implied that identifying and determining the risk scores for LUAD patients has significant prognostic relevance and

might improve the methods currently used to predict patient survival and prognosis. The risk score could also aid physicians in selecting the most appropriate course of action.

Stevens *et al*[18] reported that overexpression of the hyaluronic acid receptor HMMR in LUAD was related to molecular characteristics of inflammation and poor prognosis. In LUAD cells, decreasing HMMR expression decreased their ability to induce the development of lung carcinomas and distant metastases. In addition, HMMR is important for the orientation of the mitotic spindle in human mitotic cells and could regulate spindle assembly in mitotic cells and meiotic extracts[19]. Cell cycle analysis showed that in *PLK1*-silenced cells, the cell cycle was arrested at the G2/M phase transition. There was a positive connection between the expression levels of *PLK1* and *KPNB1* in LUAD cell lines. A decrease in *KPNB1* expression was induced by inhibition of *PLK1* and resulted in apoptosis in LUAD cells[20]. *ASPM* has been reported to be widely expressed in a variety of cancer tissues and is involved in the occurrence and progression of many malignant tumors, including lung tumors[21]. The results of Yuan *et al*[21] showed that *ASPM* depletion significantly suppressed the proliferation of lung squamous cell carcinoma cells, consistent with significant decreases in cyclin D1 and cyclin-dependent kinase 4 expression. Liu *et al*[22] reported that *ASPM* was positively related to progression and poor prognosis in prostate and hepatocellular carcinoma, but the relationship between *ASPM* and prognosis in LUAD has not been reported. Previous reports suggested that functional polymorphisms of *EXO1* might be related to the occurrence of lung cancer and that *EXO1* might be a novel biomarker for the diagnosis and treatment of lung carcinoma[23]. The KRT family is closely associated with tumor development and metastasis in various cancers, including lung, breast, and colon cancers. Han *et al*[24] suggested that *KRT16* was overexpressed in primary melanoma and metastatic melanoma and indicated significant differences in *KRT16* expression in diverse T stages and postoperative pathological stages[25]. Downregulation of *SPRR1B* suppressed the proliferation, metastasis and invasion of LUAD cells and induced G2/M arrest *in vitro*[26]. Studies have shown that hypoxia has a significant inhibitory effect on

the increase in EFNA3 expression, but there are few studies on EFNA3 in tumor cells[27]. However, hypoxia is clearly associated with prognosis and metastasis in many cancers, including LUAD[28]. All 7 genes were upregulated in LUAD tissues and were related to poor prognosis in this study.

We performed GO functional enrichment analysis based on the differentially expressed genes between the different risk groups and found that these genes were unexpectedly enriched in many processes and cytoskeletal features involved in the cell division cycle. It was rational to hypothesize that EMT might be closely related to the cell division cycle. Interestingly, there were significant differences in the enrichment of mitotic nuclear division processes between the low-risk group and the high-risk group in our research. Wang et al[29] showed that Axl inhibition was particularly synergistic with antimitotic drugs in eliminating tumor cells that had undergone EMT. In particular, tumor cells stopped dividing when they underwent EMT; thus, EMT acted as a suppressor of cancer cell division and then ended proliferation[30]. Therefore, how EMT influences tumor metastasis by affecting cell cycle processes, including mitotic nuclear division, requires further study.

Although many attempts have been made to improve the prognosis of OS, the results of these changes have not been obvious. Here, we identified 10 small molecules (GW-8510, menadione, 6-thioguanosine, phenoxybenzamine, medrysone, chlorpromazine, 8-azaguanine, meticrane, morantel and resveratrol) that have potential therapeutic benefits in OS. GW-8510 is a CDK2 inhibitor that usually inhibits the activity of cyclin/CDK complexes and negatively regulates cell cycle progression[31]. Menadione, also named VK3, is a redox cycling agent that has an antitumor effect against prostate, hepatic, lung, and breast cancer[32,33]. A previous study indicated that 6-thioguanosine, which inhibits PTHrP transcription, was used in the treatment of leukemia[34]. Phenoxybenzamine, a nonselective first-generation α1- and α2-AR antagonist, has been administered to treat lower urinary tract symptoms[35]. Medrysone is a corticosteroid usually administered to cure eye inflammation[36]. Chlorpromazine is a traditional antipsychotic or neuroleptic that is used for the treatment of schizophrenia and other

psychiatric disorders[37]. Liberante *et al*[38] showed that the concurrent targeting of RNA metabolism and splicing by 8-azaguanine offered curative prospects for *SF3B1* mutant myelodysplastic syndromes. A study found that meticrane can target the PPAR signaling pathway and inhibit the reabsorption of sodium and chloride ions in distal convoluted tubules[39]. Morantel is the classic competitive antagonist of dihydro-β-erythroidine, which noncompetitively inhibits morantel-evoked currents[40]. Resveratrol has been reported to treat a variety of cancers, including papillary thyroid cancer and liver cancer[41,42]. Overall, GW-8510, menadione, 6-thioguanosine, phenoxybenzamine, medrysone, chlorpromazine, 8-azaguanine, meticrane, morantel and resveratrol exhibit high clinical potential and warrant further study, particularly for the improvement of OS through mechanisms affecting the cell cycle.

Undoubtedly, the current research has some limitations that should be addressed. First, this was a retrospective study for the establishment of gene signatures based on public databases. Next, we must use our own esophageal cancer specimens to confirm the expression of the seven EMT-related genes in tumor tissues and adjacent tissues, as well as their correlations with clinicopathological parameters and prognosis. Second, a well-defined EMT mechanism is lacking in lung adenocarcinoma research. The biological functions of the predicted genes were estimated by computational methods, and further study is needed to reveal their mechanisms of action in tumorigenesis. In the future, EMT-related genes could be overexpressed or silenced in animal models to study tumor growth and metastasis. Additionally, we investigated how different EMT states influence tissue canceration, invasion, and metastasis in lung adenocarcinoma. Exosomes need to be studied for their role in lung adenocarcinoma initiation, development, and EMTs. Therefore, other large external multicenter experimental studies are needed to further confirm our results.

CONCLUSION

In conclusion, we constructed and validated a 7-gene prognostic risk prediction signature based on EMT-related genes in LUAD. The risk score determined by the

prognostic model of EMT-related genes was an independent risk factor for OS in LUAD. We developed a nomogram for clinicians to help predict OS in LUAD patients. According to the results of functional enrichment analysis, the pathways involved in the cell cycle and cell division were enriched in the high-risk group. In addition, we analyzed the molecular structure of potential therapeutic agents and obtained sufficient evidence indicating that EMT-related genes could be used to predict biomarkers and targeted therapies for LUAD.

ARTICLE HIGHLIGHTS

Research background

The transformation between epithelial and stromal cells during cancer progression is important in cancer progression. Understanding the relationship between epithelial-mesenchymal transition (EMT)-related genes and lung adenocarcinoma is valuable for improving its prognosis.

Research motivation

To develop an EMT-related gene signature to predict the prognosis of lung adenocarcinoma.

Research objectives

To construct and validate prognostic polygenic signatures of differentially expressed genes associated with EMT.

Research methods

We constructed a prognostic model based on differentially expressed EMT-related genes from 445 lung adenocarcinoma samples and validated its feasibility with another 439 Lung adenocarcinoma samples.

Research results

Seven EMT-related prognostic gene signatures were developed and validated. Kaplan–Meier survival analysis showed that the overall survival of patients in the high-risk group was statistically significantly poorer than that of patients in the low-risk group. The risk score based on EMT-associated genes was an independent risk factor for overall survival.

Research conclusions

The EMT-related gene signature could be used as a feasible indicator to predict the overall survival of lung adenocarcinoma patients. The molecular structures of potential therapeutic agents associated with EMT genes that could be used to treat lung adenocarcinoma were identified.

Research perspectives

More accurate genetic markers and models are needed to predict prognosis because of the low survival rate of lung adenocarcinoma.

76820_Auto_EditedC.docx

55 words — 1%

7 www.frontiersin.org
   Internet

53 words — 1%

8 Tao Zhang, Yingli Nie, Haifa Xia, Yanbin Zhang, Kailin Cai, Xiangdong Chen, Huili Li, Jiliang Wang. "Identification of Immune-Related Prognostic Genes and LncRNAs Biomarkers Associated With Osteosarcoma Microenvironment", Frontiers in Oncology, 2020
   Crossref

50 words — 1%

9 www.researchsquare.com
   Internet

42 words — 1%

10 rookieutopia.com
    Internet

39 words — 1%

11 Dandong Luo, Qiang Tao, HuiJuan Jiang, JunLing Zhu, Ning Zhang, Yongmin Zhu, Guannan Wang, YuHua Li, Yan Huang. "Identifying Eight Ferroptosis-Related Signatures for Prediction of Hepatocellular Carcinoma Overall Survival", Research Square Platform LLC, 2021
    Crossref Posted Content

37 words — 1%

12 www.dovepress.com
    Internet

37 words — 1%

13 www.science.gov
    Internet

33 words — 1%

14 Peng Zhang, Wenmiao Wang, Lei liu, HouQiang Li, XinYu Sha, Silin Wang, Zhanghao Huang, Youlang Zhou, Jiahai Shi. "Analysis of prognostic model based on immunotherapy related genes in lung adenocarcinoma", Research Square Platform LLC, 2022

32 words — 1%

15   Zhenzhen Liu, Zhentao Liu, Xin Zhou, Yongqu Lu, Yanhong Yao, Wendong Wang, Siyi Lu, Bingyan Wang, Fei Li, Wei Fu. "A glycolysis-related two-gene risk model that can effectively predict the prognosis of patients with rectal cancer", Human Genomics, 2022
Crossref

29 words — 1%

16   bmccancer.biomedcentral.com
Internet

27 words — 1%

17   Yalan Lei, Rong Tang, Jin Xu, Bo Zhang, Jiang Liu, Chen Liang, Qingcai Meng, Jie Hua, Xianjun Yu, Wei Wang, Si Shi. "Construction of a novel risk model based on the random forest algorithm to distinguish pancreatic cancers with different prognoses and immune microenvironment features", Bioengineered, 2021
Crossref

26 words — 1%

18   Guian Zhang, Yong Luo. "An Immune-Related lncRNA Signature to Predict the Biochemical Recurrence and Immune Landscape in Prostate Cancer", International Journal of General Medicine, 2021
Crossref

24 words — < 1%

19   Wen Lv, Qi Yao. "A Novel Hypoxic-Angiogenesis-Immune-Related Gene Model for Prognostic and Therapeutic Effect Prediction in Hepatocellular Carcinoma Patients", Disease Markers
Internet

23 words — < 1%

20   Min Deng, Shiquan Sun, Rongce Zhao, Renguo Guan, Zhen Zhang, Shaohua Li, Wei Wei, Rongping Guo. "The pyroptosis-related gene signature predicts prognosis and indicates immune activity in hepatocellular carcinoma", Molecular Medicine
Internet

19 words — < 1%

21 Zhuo Lin, Zhuoyan Chen, Dan Miao, Qian Xu, Liuwei Zeng, Luying Zhao, Jun Xu, Xian Song, Fujun Yu. "Pyroptosis-related genes IRAK1, NOD2, POP1 and YWHAB impact the prognosis through immune functional pathways in hepatocellular carcinoma", Research Square Platform LLC, 2021
Crossref Posted Content

19 words — < 1%

22 www.mdpi.com
Internet

19 words — < 1%

23 cyberleninka.org
Internet

17 words — < 1%

24 portlandpress.com
Internet

17 words — < 1%

25 www.spandidos-publications.com
Internet

17 words — < 1%

26 Yulin Wang, Weijiang Zhao, Zhe Xiao, Gefei Guan, Xin Liu, Minghua Zhuang. "A risk signature with four autophagy‑related genes for predicting survival of glioblastoma multiforme", Journal of Cellular and Molecular Medicine, 2020
Crossref

15 words — < 1%

27 www.nature.com
Internet

15 words — < 1%

28 Fangchao Zhao, Ren Niu, Yishuai Li, Zefang Dong, Xuebo Qin, Zhirong Li, Shujun Li. "A Stratification System of Ferroptosis and Iron-Metabolism Related Lncrnas Guides the Prediction of The Survival of Patients with Esophageal Squamous Cell Carcinoma", Research Square Platform LLC, 2021

14 words — < 1%

Crossref Posted Content

29   Lin Fang, Mengcheng Hu, Fei Xia, wenxia Bai.   14 words — < 1%
      "LINC01272 Activates Epithelial-Mesenchymal
      Transition Through miR-153-5p in Crohn's Disease", Research
      Square Platform LLC, 2021
      Crossref Posted Content

30   Shixiong Wu, Cen Zhang, Jing Xie, Shuang Li, Shuo   14 words — < 1%
      Huang. "A Five-MicroRNA Signature Predicts the
      Prognosis in Nasopharyngeal Carcinoma", Frontiers in
      Oncology, 2021
      Crossref

31   Www.Medrxiv.Org   14 words — < 1%
      Internet

32   downloads.hindawi.com   14 words — < 1%
      Internet

33   doaj.org   13 words — < 1%
      Internet

34   Ebubekir Dirican, Şeyda Tuba Savrun, İsmail   12 words — < 1%
      Erkan Aydın, Gonca Gülbay, Ülkü Karaman. "
      Analysis of mitochondrial DNA cytochrome‐b ( ) and gene
      mutations in COVID‐19 patients ", Journal of Medical Virology,
      2022
      Crossref

35   Jiayi Wang, Hongling Peng, Guangsen Zhang,   12 words — < 1%
      Yunxiao Xu, Wenzhe Yan. "The Identification and
      Analysis of a Novel Model Based on Ferroptosis-Related Genes
      for Predicting the Prognosis of Diffuse Large B-Cell
      Lymphomas", Research Square Platform LLC, 2021
      Crossref Posted Content

36 Kai Zhu, Zhichao Lang, Yating Zhan, Qiqi Tao et al. "Development and Validation of a Novel Prognostic Signature with 10 Ferroptosis-Related Genes in Acute Myeloid Leukemia", Research Square Platform LLC, 2021
Crossref Posted Content

12 words — < 1%

37 Pitroda, Sean P., Tong Zhou, Randy F. Sweis, Matthew Filippo, Edwardine Labay, Michael A. Beckett, Helena J. Mauceri, Hua Liang, Thomas E. Darga, Samantha Perakis, Sajid A. Khan, Harold G. Sutton, Wei Zhang, Nikolai N. Khodarev, Joe G. N. Garcia, and Ralph R. Weichselbaum. "Tumor Endothelial Inflammation Predicts Clinical Outcome in Diverse Human Cancers", PLoS ONE, 2012.
Crossref

12 words — < 1%

38 pdfs.semanticscholar.org
Internet

12 words — < 1%

39 Fabio Giuseppe Liberante, Katrina Lappin, Eliana M. Barros, Jekaterina Vohhodina et al. "Altered splicing and cytoplasmic levels of tRNA synthetases in SF3B1-mutant myelodysplastic syndromes as a therapeutic vulnerability", Scientific Reports, 2019
Crossref

11 words — < 1%

40 Jianfeng Ding, Xiaobo He, Wei Luo, Weiguo Zhou, Rui Chen, Guodong Cao, Bo Chen, Maoming Xiong. "Development and Validation of a Pyroptosis-Related Signature for Predicting Prognosis in Hepatocellular Carcinoma", Frontiers in Genetics, 2022
Crossref

11 words — < 1%

41 synapse.koreamed.org
Internet

11 words — < 1%

42 translational-medicine.biomedcentral.com
Internet

11 words — < 1%

43   Justus J. Bürgi, Mahendra Awale, Silvan D. Boss, Tifany Schaer et al. "Discovery of Potent Positive Allosteric Modulators of the α3β2 Nicotinic Acetylcholine Receptor by a Chemical Space Walk in ChEMBL", ACS Chemical Neuroscience, 2014
Crossref

10 words — < 1%

44   S. Chen, Y. Moroi, K. Urabe, S. Takeuchi, M. Kido, S. Hayashida, H. Uchi, T. Uenotsuchi, Y. Tu, M. Furue. "Differential expression of two new members of the p53 family, p63 and p73, in extramammary Paget's disease", Clinical and Experimental Dermatology, 2008
Crossref

10 words — < 1%

45   Weina Yu, Fengsen Liu, Qingyang Lei, Peng Wu, Li Yang, Yi Zhang. "Identification of Key Pathways and Genes Related to Immunotherapy Resistance of LUAD Based on WGCNA Analysis", Frontiers in Oncology, 2022
Crossref

10 words — < 1%

46   Xiaofeng Fu, Luwei Hong, Haiying Gong, Guangjuan Kan, Pengfei Zhang, Ting-Ting Cui, Gonglin Fan, Xing Si, Jiang Zhu. "Identification of a Nomogram with an Autophagy-Related Risk Signature for Survival Prediction in Patients with Glioma", International Journal of General Medicine, 2022
Crossref

10 words — < 1%

47   iovs.arvojournals.org
Internet

10 words — < 1%

48   link.springer.com
Internet

10 words — < 1%

49   onlinelibrary.wiley.com
Internet

10 words — < 1%

50 Di Zheng, Ling Yu, Zhun Wei, Kezhou Xia, Weichun Guo. "N6-Methyladenosine-Related lncRNAs Are Potential Prognostic Biomarkers and Correlated With Tumor Immune Microenvironment in Osteosarcoma", Frontiers in Genetics

Internet

9 words — < 1%

51 Francesco Salerno, Manuela Merli, Massimo Cazzaniga, Valentina Valeriano et al. "MELD score is better than Child–Pugh score in predicting 3-month survival of patients undergoing transjugular intrahepatic portosystemic shunt", Journal of Hepatology, 2002

Crossref

9 words — < 1%

52 Jingxin He, Xinwei Jiao, Xin Sun, Yijia Huang, Pengyang Xu, Yunxia Xue, Ting Fu, Jun Liu, Zhijie Li. "Short-Term High Fructose Intake Impairs Diurnal Oscillations in the Murine Cornea", Investigative Opthalmology & Visual Science, 2021

Crossref

9 words — < 1%

53 Weiyu Dai, Yizhi Xiao, Weimei Tang, Jiaying Li et al. "Identification of an EMT-Related Gene Signature for Predicting Overall Survival in Gastric Cancer", Frontiers in Genetics, 2021

Crossref

9 words — < 1%

54 Yiqun Jin, Bai. Xue-song. "A Novel Pyroptosis-related Genes Prognostic Signature Affects the Immune Infiltration Patterns in Hepatocellular Carcinoma", Research Square Platform LLC, 2021

Crossref Posted Content

9 words — < 1%

55 Zhihong Cai, Yijing Cao, Yichen Luo, Haobin Hu, Hui Ling. "Signalling mechanism(s) of epithelial–mesenchymal transition and cancer stem cells in tumour therapeutic resistance", Clinica Chimica Acta, 2018

9 words — < 1%

Crossref

56 Zhixiang Yu, Haiyan He, Yanan Chen, Qiuhe Ji, Min
    Sun. "A novel ferroptosis related gene signature is
    associated with prognosis in patients with ovarian serous
    cystadenocarcinoma", Scientific Reports, 2021
    Crossref

9 words — < 1%

57 Zhiyong Zeng, Junfang Lin, Kejie Zhang, Xizhe Guo,
    Xiaoqiang Zheng, Apeng Yang, Junmin Chen.
    "Single cell RNA-seq data and bulk gene profiles reveal a novel
    signature of disease progression in multiple myeloma", Cancer
    Cell International, 2021
    Crossref

9 words — < 1%

58 wjso.biomedcentral.com
    Internet

9 words — < 1%

59 www.jci.org
    Internet

9 words — < 1%

EXCLUDE QUOTES   OFF   EXCLUDE SOURCES   OFF
EXCLUDE BIBLIOGRAPHY   OFF   EXCLUDE MATCHES   < 9 WORDS