IDENTIFICATION OF NOVEL EPITHELIAL-MESENCHYMAL TRANSITION (EMT) AND DNA DAMAGE REPAIR (DDR)-RELATED PROGNOSTIC GENE SIGNATURES AS PREDICTIVE BIOMARKERS IN SIX CANCER COHORTS

ABSTRACT

Background: Breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pancreatic adenocarcinoma (PAAD), and stomach adenocarcinoma (STAD) account for ~40.97% of all cancer cases and ~46.35% of cancer-related deaths worldwide. Biomarker limitations restrict therapy options, making these cancers difficult to cure. Here, we present a novel dual-signature, multi-cancer comparative analysis integrating epithelial-mesenchymal transition (EMT) and DNA damage repair (DDR) prognostic gene signatures.

Methods: RNA-Seq profiles and corresponding clinical data from The Cancer Genome Atlas (TCGA) project were used to identify differentially expressed genes (DEG). Univariate Cox, Kaplan-Meier (KM), and Least Absolute Shrinkage and Selection Operator (LASSO) regression analyses were applied to identify prognostic EMT- and DDR-related genes and construct final models. Validation was performed using univariate and multivariate Cox, ROC curves, AUC analysis, t-SNE, and Principal Component Analysis (PCA). Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Gene Ontology-Biological Process (GO-BP) were performed to assess pathways involved in the prognostic signatures. CIBERSORT.X and ESTIMATE algorithms were used to analyse immune infiltration, and a nomogram model was developed by integrating clinical factors to identify prognostic signatures. Drug sensitivity correlation analysis was performed between the risk score of cancer cell lines and the IC_{50} values of drugs. The prognostic ability of identified EMT- and DDR-related signatures was validated in Gene Expression Omnibus (GEO) datasets.

Results: Based on Cox, KM, and LASSO analysis, we identified a novel prognostic six-gene related signature in BRCA, six- and two-gene in COAD, nine- and five-gene in LUAD, four-gene in LUSC, ten-, and eight-gene in PAAD, and nine-, and seven-gene related signatures in STAD for EMT and DDR, respectively. ROC and AUC values indicated effective performance

across cancers, particularly in PAAD. PCA and t-SNE confirmed strong discrimination at the transcriptomic level. UniCox and MultiCox analyses demonstrated that EMT/DDR-related signatures-based risk scores were independent prognostic variables. Time-dependent ROC analysis showed that the nomogram achieved higher AUCs than individual clinical factors. Immune analysis revealed reduced infiltration of macrophages (M0) and naïve B cells in high-risk group suggesting an immunosuppressive tumour microenvironment in high-risk patients. Correlation analysis with drug IC_{50} values in cancer cell lines suggested that high-risk scores were associated with predicted sensitivity to (e.g., DNA-damaging agents) in BRCA and COAD, but with predicted resistance to (e.g., certain chemotherapies) in the other cancer types.

Conclusion: The proposed EMT- and DDR-related prognostic gene signatures performed well for overall survival prediction across cancers and could serve as promising biomarkers.

ABSTRACT

Background: Il carcinoma mammario invasivo (BRCA), l'adenocarcinoma del colon (COAD), l'adenocarcinoma polmonare (LUAD), il carcinoma polmonare a cellule squamose (LUSC), l'adenocarcinoma pancreatico (PAAD) e l'adenocarcinoma gastrico (STAD) rappresentano circa il 40,97% di tutti i casi di tumore e circa il 46,35% dei decessi correlati al cancro a livello mondiale. Le limitazioni dei biomarcatori restringono le opzioni terapeutiche, rendendo questi tumori difficili da curare. In questo studio presentiamo una nuova analisi comparativa multi tumore basata su una doppia firma, che integra firme geniche prognostiche relative alla transizione epitelio mesenchimale (EMT) e alla riparazione del danno al DNA (DDR).

Methods: I profili di RNA Seq e i corrispondenti dati clinici del progetto The Cancer Genome Atlas (TCGA) sono stati utilizzati per identificare i geni differenzialmente espressi (DEG). Analisi di Cox univariata, Kaplan Meier (KM) e regressione LASSO (Least Absolute Shrinkage and Selection Operator) sono state applicate per identificare i geni prognostici correlati a EMT e DDR e per costruire i modelli finali. La validazione è stata eseguita mediante analisi di Cox univariata e multivariata, curve ROC, analisi delle AUC, t SNE e analisi delle componenti principali (PCA). Le analisi basate sulla Kyoto Encyclopaedia of Genes and Genomes (KEGG) e sulla Gene Ontology Biological Process (GO BP) sono state utilizzate per valutare i pathway coinvolti nelle firme prognostiche. Gli algoritmi CIBERSORT.X ed ESTIMATE sono stati impiegati per analizzare l'infiltrazione immunitaria, e un nomogramma è stato sviluppato integrando fattori clinici per definire le firme prognostiche. È stata inoltre eseguita un'analisi di correlazione tra la sensibilità ai farmaci e il punteggio di rischio delle linee cellulari tumorali e i valori di IC_{50} dei farmaci. La capacità prognostica delle firme correlate a EMT e DDR identificate è stata validata in dataset del Gene Expression Omnibus (GEO).

Results: Sulla base delle analisi di Cox, KM e LASSO, abbiamo identificato una nuova firma prognostica a sei geni in BRCA, firme a sei e due geni in COAD, a nove e cinque geni in LUAD, a quattro geni in LUSC, a dieci e otto geni in PAAD e a nove e sette geni in STAD per EMT e DDR rispettivamente. I valori di ROC e AUC hanno indicato una performance efficace in tutti i tumori, in particolare in PAAD. Le analisi PCA e t SNE hanno confermato una marcata capacità di discriminazione a livello trascrittomico. Le analisi UniCox e MultiCox hanno dimostrato che i punteggi di rischio basati sulle firme correlate a EMT e DDR rappresentano variabili prognostiche indipendenti. L'analisi ROC dipendente dal tempo ha mostrato che il nomogramma raggiungeva valori di AUC superiori rispetto ai singoli fattori clinici. L'analisi

immunologica ha rivelato una ridotta infiltrazione di macrofagi M0 e di linfociti B naive nel gruppo ad alto rischio, suggerendo un microambiente tumorale immunosoppressivo nei pazienti ad alto rischio. L'analisi di correlazione con i valori di IC_{50} dei farmaci nelle linee cellulari tumorali ha suggerito che punteggi di rischio elevati erano associati a una sensibilità prevista, per esempio ad agenti che danneggiano il DNA, in BRCA e COAD, ma a una resistenza prevista, per esempio ad alcune chemioterapie, negli altri tipi di tumore.

Conclusion: Le firme geniche prognostiche correlate a EMT e DDR proposte mostrano buone prestazioni nella predizione della sopravvivenza globale in diversi tipi di tumore e potrebbero rappresentare promettenti biomarcatori.