Comprehensive Analysis of a Cuproptosis-Related Gene Index in Pancreatic Adenocarcinoma with Significant Implications on Prognosis and Immunotherapy as well as Chemotherapy Efficacy

Xufeng Huang
University of Debrecen
Shujing Zhou
University of Debrecen
Ying Liu
Sixth Medical Center of PLA General Hospital
János Tóth
University of Debrecen
Hajdu András (hajdu.andras@inf.unideb.hu)
University of Debrecen

Research Article

Keywords: Cuproptosis, machine learning, pancreatic cancer, gene signature, tumor microenvironment, cancer immunity cycle, immunotherapy, chemotherapy

Posted Date: June 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1746948/v1

Abstract

Background

A novel type of cell death induced by intracellular copper accumulation, principally different from other known ones such as autophagy, ferroptosis, and pyroptosis, was discovered and referred to as “Cuproptosis” recently. As the underlying mechanism was newly identified, its potential connection to pancreatic adenocarcinoma (PAAD) is still an open issue. However, interestingly, one of the 10 Cuproptosis-related genes, CDKN2A, was previously reported as a driver mutation gene in pancreatic adenocarcinoma. Therefore, it raised our interest to develop a Cuproptosis-related gene index (CRGI) for better overall survival (OS) estimation, explore the corresponding relationship with tumor immunology, and discuss the chemotherapy efficacy of commonly used anti-cancerous drugs relevant to CRGI.

Methods

Gene expression profiles and matched clinical data were curated from publicly accessible sources from which the CRGI was developed. Immunogenic properties were investigated according to the immune checkpoints, potential response to the immunotherapy, cancer immunity cycle, etc. The sensitivity to chemotherapy agents was evaluated through the IC50 curve. The consensus clustering approach was implemented to robustly identify the CRGI-based molecular subtypes and explore the immunotherapy and chemotherapy efficacies.

Results

A novel gene signature derived from Cuproptosis was established. Compared with those signatures originating from other cell death mechanisms, it exerted a more robust yet more accurate predictive performance. Besides, the result of correlation analysis unraveled the distinct differences in tumor immunology between the high- and low-risk groups. Through in-depth analysis of the IC50 curve of each drug, the predictive chemotherapy efficacy of 32 commonly used anti-cancerous drugs was also found various in high- and low-risk groups. The analysis of molecular subtypes has further confirmed these ideas.

Conclusion

In the present study, we highlighted the outstanding achievement of CRGI in PAAD prognostic prediction and its association with tumor immunology. These findings may benefit future immunotherapy- and chemotherapy-based interventions.

Introduction

Cuproptosis as a novel cell death mechanism has been raising great attention in the scientific community[1]. Since the other types of cell death such as autophagy, ferroptosis, and pyroptosis were proven to be closely associated with tumorigenesis, metastasis, and prognosis[2,3,4], we were curious about the potential connection between Cuproptosis and various cancers, particularly pancreatic adenocarcinoma (PAAD), due to its remarkably poor survival rate in 5-year overall survival (OS)[5,6,7,8,9], and the interesting fact that one of the 10 Cuproptosis-related genes, CDKN2A, was a deemed to be a driver mutation gene in PAAD[10,11,12,13], implicating a bridge that may link Cuproptosis to PAAD. Therefore, in the present study, through 10 mainstream machine learning methods, we optimized a Cuproptosis-gene index (CRGI) with an important role in prognosis, tumor immunology, the classification of molecular subtypes, and the efficacy of immunotherapy and chemotherapy.

Figure 1 demonstrates the workflow of the present study briefly.

Figure 1 Graphical abstract of the present study.

Methods

Collection and integration of the transcriptome data and matched clinical information

In the present study, we retrospectively curated 10 Cuproptosis-related genes from the latest reports[1] as well as the transcriptomic data and matched clinical information on pancreatic adenocarcinoma (PAAD) from publicly accessible sources, including the Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga, N = 177) and the Australian dataset of International Cancer Genome Consortium (ICGC, https://www.icgc-argo.org, N = 269)[14]. All the data involved in the present study were processed by R (v 4.0.3) and corresponding R packages. Notably, if it wasn't specifically mentioned, P-value<0.05 is considered statistically significant and might be annotated as * within the figures. Moreover, **, ***, and **** might appear within the figures to indicate the P-value thresholds 0.01, 0.001, and 0.0001, respectively.

Machine learning in the development of Cuproptosis-related gene index (CRGI)

10 mainstream machine learning algorithms were used in the optimization of CRGI, including least absolute shrinkage and selection operator (LASSO), decision tree, Gaussian mixture model, gradient boosted decision trees (GBDT), K-nearest neighbors (K-NN), light gradient boosting machine (LGBM), logistic regression, random forest, support vector machine (SVM), and extreme gradient boosting (XGboost)[15,16,17,18,19,20,21,22,23]. Their performances were assessed by the time-dependent receiver operative characteristic (ROC) curves in which the area under the curve (AUC) represented the predictive power. The greater the AUC value indicated the better accuracy and robustness of the model. The ROC curve was created by the R package "timeROC".

Decision curve analysis (DCA)

Usually, prognostic models and diagnostic tests are mathematically evaluated with measures of accuracy that do not consider clinical outcomes. To help improve such shortcomings, DCA was developed[24]. It is often used to compare the efficacy of different predictive models and diagnostic tests to maximize the clinical benefits when false positives and false negatives are known to be unavoidable. In the present study, we performed this analysis by using the R package “ggDCA”.

Construction of a conventional nomogram and corresponding calibration curve

The risk score of our model that was optimized through the machine learning method as mentioned was integrated as a prognostic indicator with other clinical factors to estimate the overall survival (OS) probability, through univariate and multivariate Cox regression, and a traditional nomogram with calibration curve were constructed from these results. The visualization was achieved by using the R packages “forestplot”, and “rms”[25,26].

Construction of an online OS calculator

By utilizing the analytic results acquired from the last step, we built an easy-to-use web-based OS calculator by the R package “DynNom”[27].

The calculator is available at https://debmed.shinyapps.io/CRGIProgCal/

Gene set enrichment analysis (GSEA)

The GSEA software (v 4.0.3, http://software.broadinstitute.org/gsea/index.jsp) was used to investigate the underlying mechanism of the high- and low-risk groups using the GO terms and KEGG pathways databases[28,29,30,31,32,33].

Estimation of the tumor microenvironment condition

The tumor microenvironment condition was estimated quantitatively by using the R package “ESTIMATE” through a scoring system, including stromal score, immune score, and ESTIMATE score based on the mRNA expression[34].

Screening of immune cell infiltration

The gene expression profiles were processed by integrating 7 mainstream immunoinformatic algorithms, including TIMER, CIBERSORT, CIBERSORT–ABS, QUANTISEQ, MCPCOUNTER, XCELL, and EPIC, and the immune cell infiltration matrix was obtained[35,36,37]. The R package “ggplot2” was used to visualize the distribution of infiltration of diverse immune cell types as a heatmap.

Prediction of the potential immune checkpoint blockade response

The concept of immunotherapy for tumors was proposed at the end of the 19th century and refers to a treatment method that uses the body's immune system to destroy cancer cells. The therapies that use immune checkpoint blockade have revolutionized the treatment of human cancer. Herein, the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was applied to predict the responsiveness of a single sample or a subtype based on expression profiling data.

Unsupervised clustering through the K-means algorithm

The consistency analysis was performed using the R package “ConsensusClusterPlus (v 1.54.0)”, the maximal number of clusters was 6, and 80% of the samples were extracted 100 times through a re-sampling approach[38]. The package generated the consensus matrix, empirical cumulative distribution function (CDF), and delta area plots for each selected K value.

Moreover, as a complementary confirmation, a principal component analysis (PCA) was conducted to elucidate if the samples were well-separated with the batch effect fully removed.

Survival analysis

The Kaplan Meiler curve was applied to compare the survival difference between different groups. The P-value and hazard ratio (HR) with a 95% confidence interval (CI) were generated by log-rank test and univariate Cox proportional hazards regression. Both were done by the R package “survival”.

In-silica evaluation of chemosensitivity

32 commonly used anti-cancerous drugs were involved in the present study. The IC50 values of these drugs were predicted from the expression matrix by the pRRopheticPredict function in the R package

“pRRophetic”[39].

Results

CRGI was optimized from 10 mainstream algorithms

10 Cuproptosis-related genes were curated from the work of Tsvetkov et al[1]. Combined with 6 well-recognized driver mutation genes (i.e., KRAS, TP53, SMAD4, BRCA1, BRCA2, and CDKN2A) in PAAD, they were subjected to mainstream machine learning procedures to develop a Cuproptosis-related gene index (CRGI)[40].

In the TCGA dataset, among 10 mainstream machine learning algorithms, we optimized the best model through LASSO penalized Cox regression that had a leading AUC value in 1, 2, 3, and 4-year overall survival (OS) predictive performance, up to 0.736, 0.703, 0.708, 0.812, respectively (Figure 2 C). The formula for the risk score calculation is:

0.5316*KRAS + 0.014*TP53 – 0.0407*CDKN2A – 0.0999*SMAD4 + 0.3768*BRCA1 + 0.0866*BRCA2 – 0.1292*LIAS – 0.587*LIPT1 – 0.3158*DLD + 0.4833*DLAT – 0.3627*PDHA1 – 0.3253*MTF1 – 0.1286*GLS

Following the calculation of the risk score, patients were separated into high- and low-risk groups, and apparently, there was a higher mortality rate within the high-risk group relative to that of the low-risk group (Figure 2 A). The survival analysis further revealed the fact that the low-risk group possessed a significant survival advantage (Figure 2 B). The difference in expression of the CRGI genes between high- and low-risk groups was visualized in the manner of a heatmap as shown at the bottom of Figure 2 A.

Interestingly, although the predictive performances of the other machine learning algorithms in the first 4 years were poor, 7 algorithms (i.e., Decision Tree, Gaussian, K-NN, Logistic Regression, Random Forest, SVM, and XGBoost) possessed an AUC value = 1.000 in 5-year OS prediction (Figure 2 D-L). Taking it all together, we decided to choose LASSO as the final predictive model for the following comprehensive analytics.

Figure 2 Predictive performance comparison of 10 mainstream machine learning algorithms. (A) Distribution of the risk score and survival status in high- and low-risk groups, and the heatmap that depicts the expression of CRGI genes between high- and low-risk groups. (B) Survival analysis of high- and low-risk groups. (C) The time-dependent ROC curve with the AUC value of 1-, 2-, 3-, 4-, and 5-year OS prediction of the best model optimized by LASSO penalized Cox regression. (D)-(L) The predictive accuracy of other machine learning algorithms (i.e., Decision Tree, Gaussian Mixture Model, GBDT, K-NN, LGBM, Logistic Regression, Random Forest, SVM, and XGboost, respectively).

The predictive performance of CRGI was superior to that of the signatures derived from other cell death mechanisms

Recently, with the progressions in the in-depth understanding of cell death mechanisms, a considerable number of prognosis-predictive gene signatures have been proposed. To clarify whether CRGI behaves better than those signatures originating from other underlying mechanisms, we retrieved signatures that were derived from autophagy, ferroptosis, and pyroptosis gene set[41,42,43].

We performed time-dependent receiver operative characteristic (ROC) curves across the TCGA (Figure 2 C, Figure 3 A-C) and ICGC datasets (Figure 3 D-G) for each signature and observed that our model was significantly associated with prognosis in both cohorts and possessed the best AUC values and decision curves, demonstrating the stability and accuracy of CRGI.

Usually, it was considered rigorous enough to assess different models by comparing the AUC values of the ROC curve. However, as ROC analysis merely accounts for the specificity and sensitivity of the model, in the field of medicine, in case of unavoidable false positives and false negatives, one should maximize the clinical benefits from either result as possible. Therefore, we complementarily employed decision curve analysis (DCA) for each signature in the TCGA and ICGC datasets, the result of which again supported that our model had the most advanced predictive power (Figure 3 H-I).

Figure 3 Comparison of other cell death mechanisms-based prognostic signatures in PAAD. (A)-(C) The ROC curve of autophagy-, ferroptosis-, and pyroptosis-based models in the TCGA dataset, respectively. (D)-(G) The ROC curve of our model, and the autophagy-, ferroptosis-, and pyroptosis-based ones in the ICGC dataset, respectively. (H) DCA of our model, and the autophagy-, ferroptosis-, and pyroptosis-based ones in the TCGA dataset, respectively. (I) DCA of our model, and the autophagy-, ferroptosis-, and pyroptosis-based ones in the ICGC dataset, respectively.

CRGI served as an independent and reliable indicator in PAAD prognostic prediction

Based on our model, we extract the risk score, 13 CRGI genes (i.e., KRAS, TP53, SMAD4, BRCA1, BRCA2, CDKN2A, DLAT, LIPT1, LIAS, DLD, PDHA1, MTF1, GLS), age, gender, pathological status, TNM stages, histological grades, radiotherapy, smoking, and drinking information from the TCGA dataset, and carried out univariate Cox regression to examine if they are statistically correlated with prognosis and multivariate Cox regression to qualify their eligibility as independent prognostic indicators. It turned out that the risk score, KRAS, SMAD4, BRCA2, LIAS, age, pathological T stage, pathological N stage, and radiotherapy were associated with prognosis as a result of the univariate Cox regression, while results of multivariate Cox regression furtherly indicated that the risk score, LIAS, and smoking were independent prognostic indicators (Figure 4 A-B).

According to the mentioned findings, we constructed a conventional nomogram and its calibration curve (Figure 4 C-D) which suggested that our model accurately predicted the prognosis of PAAD with a C-index value up to 0.795 (95% CI: 0.743-0.846, P-value = $2.78e^{-29}$).

To make it more user-friendly, the underlying statistics were implemented in a web-based OS calculator which assisted the clinicians to estimate the OS probability by entering the clinicopathological parameters, and the survival time of interest for prediction (Figure 4 E). The calculator is available at https://debmed.shinyapps.io/CRGIProgCal/. The clinical data behind the calculator is available in the supplementary material T1_Rare data for the web-based OS calculator.

We also explored the correlation between the risk score and different clinical factors (e.g., age, gender, tumor grade, etc.). It appeared that the risk score was only associated with tumor grade, between G1 and G3, G2 and G3 (S1_Correlation between the risk score and clinical factors).

Figure 4 Nomograms with clinicopathological characteristics to predict OS in PAAD. (A)-(B) Forest plots of the results of the univariate and multivariate Cox regression, respectively. (C)-(D) Conventional nomogram and its calibration curve based on our model. (E) The screenshot of the web-based OS calculator that was developed from the underlying statistics and calculating a fictional case.

Over-representative analysis revealed the functional importance of CRGI in PAAD

Over-representative analysis was conducted to unravel the functional mechanisms underlying high- and low-risk groups through the ssGSEA. The result of correlation analysis of the risk score and ssGSEA score implied that they were statistically and positively associated, with a P-value = $8.15e^{-03}$ and Spearman coefficient = 0.2 (Figure 5 A). Moreover, different cancerous hallmarks were found enriched in high- and low-risk groups with statistical significance. Consequently, 34 pathways were identified within the high-risk group. The most enriched pathways were upregulated and mostly related to cell proliferation, including G2M CHECKPOINT, E2F_TARGETS, etc. (Figure 5 B). Within the low-risk group, 4 pathways were downregulated, mainly related to digestive functions, such as BILE_ACID_METABOLISM (Figure 5 C).

We also performed GO terms and KEGG enrichment analysis for the CRGI genes. Subsequently, GO terms were found most relevant to the energy production within the mitochondria (Figure 5 D). In particular, KEGG pathways were found directly connected to pancreatic cancer (Figure 5 E).

Figure 5 Over-representative analysis of CRGI in PAAD. (A) Correlation analysis of the risk score and ssGSEA score. (B) The top 10 most enriched and upregulated cancerous hallmarks in the high-risk group.

(C) The 4 downregulated cancerous hallmarks in the low-risk group. (D) The enriched GO terms of CRGI genes. (E) The enriched KEGG pathways of CRGI genes.

CRGI was associated with the tumor microenvironment condition, cancer immunity cycle, and immunotherapy efficacy

It has been widely believed that cancers are essentially considered as dynamic ecosystems wherein subclone populations of most cancer cells and non-malignant cells in the tumor microenvironment engage cooperatively to promote the disease progression. Therefore, it is necessary to investigate the tumor microenvironment condition. Herein, we utilized the R package “ESTIMATE” to elucidate it in a quantitative way, through which we found that except for the stromal score, the immune score and ESTIMATE score were found statistically significant and that higher immune score, as well as higher ESTIMATE score, were observed in normal tissues (Figure 6 A-C). Then, we analyzed the correlation between the risk scores and stromal, immune, and ESTIMATE scores, respectively. It was found that the risk score was negatively associated with the immune score with statistical significance (Figure 6 D-F). We also analyzed the mentioned scores in the high- and low-risk groups, and it was observed that there was a difference in immune score and ESTIMATE score with statistical significance where the lower-risk group possessed a relatively higher immune score and ESTIMATE score (Figure 6 G). These findings supported the idea that CRGI as a classifier of risk group played an important role in the distinguishment of the tumor microenvironment condition in PAAD. On the other hand, we exhaustively screened the immune cell types in the tumor infiltration process in different risk groups by integrating 7 mainstream immunoinformatic algorithms, including TIMER, CIBERSORT, CIBERSORT–ABS, QUANTISEQ, MCPCOUNTER, XCELL, and EPIC. It was found that the immune cell types involved in this process were very much diverse (Figure 6 H).

Immune checkpoints are negatively regulatory proteins in the immune system that are indispensable for the maintenance of self-tolerance, the prevention of autoimmune responses, and the minimalization of tissue damage. They function by controlling the timing and intensity of immune response. In immunotherapy, the overexpression of immune checkpoints inhibits the function of immune cells so that the body cannot produce an effective anti-tumorous immune response, which ultimately leads to immune escape. Therefore, to fully evaluate the potential impact on immunotherapy efficacy caused by CRGI, we also analyzed the difference in the expression of the representative genes of the soundest immune checkpoints in high- and low-risk groups, including PD-1 (i.e., protein of PDCD1 gene), PD-L1 (i.e., protein of CD274 gene), PD-L2 (i.e., protein of PDCD1LG2 gene), CTLA4, HAVCR2, LAG3, TIGIT, and SIGLEC15. It was found that there are statistical significances in the expression of PD-1, CTLA4, TIGIT, and LAG3 (Figure 6 I). It was observed that there were relatively lower expression levels in the low-risk group, which indicated a betterment in immunotherapy efficacy. In addition, after integrating the data curated from the TCGA cohort, by utilizing the TIDE algorithm (i.e., a scoring system that inversely reflects the immunotherapy efficacy), we predicted the potential response of PAAD patients to immune checkpoint

blockade. To be more exact, we first carried out a correlation analysis in which we found that the TIDE score and the risk score showed a significantly positive correlation with a P-value = 0.01 and Spearman coefficient = 0.19 (Figure 6 J). Then, we stepped forward, finding that in high- and low-risk groups, there was also a distinct difference in which the low-risk group demonstrated a prominently lower TIDE score than that of the high-risk group, implying that there was an advancement in immunotherapy efficacy in the low-risk group (Figure 6 K). On the other hand, tumor mutation burden (TMB) and Microsatellite instability (MSI) as other reliable references for the prediction of immunotherapy efficacy are also worth being studied. Therefore, we extensively analyzed their relationship with the risk score. It was found that there was statistical significance regarding TMB, but not for MSI (Figure 6 L, S2_Correlation with MSI).

As Cuproptosis is a cell death mechanism that may raise immune reactions, it is of great interest to investigate the underlying mechanism in the cancer immunity cycle[44] (Figure 6 M). Correlation analysis between the risk score and cancer–immunity cycle as well as known biological functions in pancreatic cancer showed that the risk score presented a significantly positive correlation to the processes of cancer antigen presentation, recognition of cancer cells by T cells, killing of cancer cells in cancer immunity cycle, and tumor proliferation signature, cellular response to hypoxia, EMT markers, apoptosis, DNA repair, DNA replication, G2M checkpoint, PI3K AKT mTOR pathway, MYC targets, P53 pathway, TGFB, collagen formation, and ferroptosis in known biological functions (Figure 6 N). Similar differences with statistical significance existed between high- and low-risk groups, indicative of the remarkable interactions of CRGI with tumor immunology (Figure 6 O).

Figure 6 In-depth analytics on the relationship between CRGI and the tumor microenvironment condition, immune cell infiltration, immunotherapy efficacy, TMB, as well as the cancer immunity cycle. (A)-(C) Comparison of the stromal, immune, and ESTIMATE scores of tumorous and normal tissues in the TCGA dataset. (D)-(F) Correlation analysis between the risk score and the stromal, immune, and ESTIMATE scores. (G) The violent plot demonstrated the comparison of the stromal, immune, and ESTIMATE scores in high- and low-risk groups. (H) The heatmap demonstrated the diverse immune cell types in the infiltration process. (I) The violent plot demonstrated the comparison of the expression of the representative genes of the soundest immune checkpoints in high- and low-risk groups. (J) Correlation analysis between the risk score and TIDE score. (K) The violent plot demonstrated the comparison of the TIDE scores in high- and low-risk groups. (L) Correlation analysis between the risk score and TMB score. (M) Graphical demonstration of the cancer immunity cycle. (N) Correlation analysis between the risk score and the main steps of the cancer immunity cycle as well as known biological processes. (O) The boxplot demonstrated the comparison of marker scores in high- and low-risk groups.

CRGI-based risk groups possessed various chemosensitivity

Chemotherapy has been the centerpiece in the treatment of cancer over the past few decades, yet due to the heterogeneous characteristics of tumors, even the responses to the same chemotherapeutic may vary from one patient to another[45]. To address this problem, genome-based methodologies must be introduced. For this purpose, we evaluated the chemosensitivity of PAAD patients from the TCGA dataset who were classified into high- and low-risk groups previously in the present study to 32 commonly used anti-cancerous drugs (S3_Comparison of the efficacy of 32 anti-cancerous drugs in high- and low-risk groups). Subsequently, the IC50 value of 5 drugs (i.e., Lenalidomide, Metformin, Temsirolimus, Axitinib, and Camptothecin) was found relatively higher in the high-risk group (Figure 7 A), while that of the other 11 drugs (i.e., Paclitaxel, Lapatinib, Dasatinib, Bleomycin, Docetaxel, Doxorubicin, Bexarotene, Gefitinib, Bosutinib, and Bortezomib) was found relatively higher in the low-risk group (Figure 7 B).

Figure 7 Comparison of the efficacy of 32 chemotherapeutics in (A) high-risk group, and (B) low-risk group.

CRGI-based molecular subtypes were characterized by different survival outcomes and immunotherapy efficacy

Since the above analytics has revealed that high- and low-risk groups possessed distinct OS probability and immunotherapy efficacy, it raised our interest in systemically dividing it into more precise molecular subtypes through an unsupervised consensus method (i.e., K-mean algorithm). It was found that when K = 4, the PAAD samples were separated into 4 clusters in the consensus diagram (Figure 8 A). Meanwhile, when K = 4, CDF almost reached its maximum which indicated good stability (Figure 8 B). Besides, it was observed that CDF changed only slightly when K+/-1 (Figure 8 C). Therefore, K = 4 was deemed to be an ideal value in the present study. To ensure the robustness of the clustering, we also conducted a principal component analysis (PCA), through which we could see that the samples were indeed well separated (Figure 8 D). Therefore, ultimately, 4 molecular subtypes were identified, annotated by C1 (N = 37), C2 (N = 90), C3 (N = 23), and C4 (N = 29).

Furthermore, we examined the expression profiles of the 13 CRGI genes in each molecular subtype and found that generally, the expression levels were arranged in such order: C1 > C4 > C2 > C3 (Figure 8E).

We then investigated the clinical outcomes in these molecular subtypes. Results of the survival analysis suggested that C2 had the best prognosis, followed by C4, C3, and C1 (Figure 8 F).

Finally, we inspected the TIDE score and the expression of the representative genes of the soundest immune checkpoints in different molecular subtypes. For the TIDE score, there were differences with statistical significance between C1 and C3, and between C2 and C3, where C3 had a relatively lower TIDE score compared with that of C1 and C2 (Figure 8 G). For the expression of the representative genes,

except SIGLEC15, the rest were all found statistically significant and exhibited prominent differences in each molecular subtype (Figure 8 H). In short, the 4 molecular subtypes demonstrated diverse immunogenic features, and that may lead to various efficacy in immunotherapy.

Figure 8 Characterization of 4 CRGI-based molecular subtypes with different survival outcomes and immunotherapy efficacy. (A) The consensus matrix (K= 5), where the columns and rows represented the TCGA samples involved which were clustered into squares in white to deep blue colors. (B) The consensus CDF plot (K = 1, 2, 3, 4, 5, 6). (C) The Delta area plot (K = 1, 2, 3, 4, 5, 6). (D) The PCA diagram where TCGA samples were represented in the manner of scattering spots. The circle bordered the area within which all the samples belonged to the same molecular subtype. (E) The boxplot demonstrated the expression profiles of the 13 CRGI genes in each molecular subtype. (F) Survival analysis demonstrated the clinical outcomes of different molecular subtypes. (G) The boxplot demonstrated the comparison of the TIDE scores in different molecular subtypes. (H) The boxplot demonstrated the expression profiles of the representative genes of the soundest immune checkpoints in different molecular subtypes.

CRGI-based molecular subtypes also possessed various chemosensitivity

Like what we have done to the high- and low-risk groups, we carried out the elucidation of the chemosensitivity of PAAD patients from the TCGA dataset who were classified into 4 molecular subtypes according to the CRGI. As a result, we found that 27 out of 32 commonly used anti-cancerous drugs exerted statistically significant differences in the 4 molecular subtypes (Figure 9). Notably, Axitinib was probably the most distinct one as it differed between every two molecular subtypes with extremely high statistical significance.

Figure 9 Comparison of the efficacy of 32 chemotherapeutics in 4 CRGI-based molecular subtypes.

Discussion

Thanks to the modern methodologies innovated in basic medicine and the advancements achieved in clinical practice, the long-term overall survival (OS) rates of most cancer types have been much improved. However, pancreatic adenocarcinoma (PAAD) remains the most devastating disease with its 5-year OS rate maintained in the single digits[2,3,4]. One even more crucial fact is that, due to the complexity of tumor biology, it is insufficient to predict the survival outcomes through merely clinical and pathological features or a sole biomarker, and thus effective estimation of the OS rate in PAAD has been another concern. Therefore, in the present study, we combined machine learning approaches with a novel cell

death mechanism, Cuproptosis, to optimize a Cuproptosis-related gene index (CRGI) to contribute to this situation.

To the best of our knowledge, this is the first study that investigated prognostic prediction from the perspective of Cuproptosis in PAAD. The index was composed of 13 genes including well-verified biomarkers such as KRAS, TP53, SMAD4, BRCA1, BRCA2, and CDKN2A, as well as newly discovered Cuproptosis-related genes including DLAT, LIPT1, LIAS, DLD, PDHA1, MTF1, and GLS[40]. Interestingly, among these genes, CDKN2A was a well-known driver mutation gene in PAAD and a member of the 10 Cuproptosis-related genes. Other CRGI genes were essential components in the generation process of mitochondrial energy. For example, DLAT was an important part of the pyruvate dehydrogenase complex that was deeply involved in the clump-up of lipoylated TCA cycle proteins in Cuproptosis[46].

Compared with other signatures derived from autophagy, ferroptosis, and pyroptosis, our model exerted a more robust but also more accurate predictive performance. Notably, this short conclusion was not merely drawn from conventional assessment (i.e., through the area under the curve in the time-dependent receiver operative characteristic analysis), but also after the consideration of clinical benefits calculated from decision curve analysis (DCA) as we fully understand that it is necessary to maximize the goodness in clinical practice when reasonable false positivity and false negativity were unavoidable[24].

In addition, we explored the correlation between the risk score and tumor immunology as well as the differences in tumor immunology between the high- and low-risk groups classified by CRGI, through which we found that CRGI was associated with the tumor microenvironment condition, cancer immunity cycle, and immunotherapy efficacy. Regarding immunotherapy particularly, the lower-risk group was advantageous as it possessed a relatively better response to immune checkpoints blockade and was more closely associated with the cancer immunity cycle as well as some known biological processes in pancreatic cancer.

The predictive chemotherapy efficacy of 32 commonly used anti-cancerous drugs was screened and found various in IC50 curves which quantitatively described the efficacy of a particular inhibitory substance (i.e., the anti-cancerous drug) that is needed to block half of a given biological process or component. Among them, the IC50 value of Lenalidomide, Metformin, Temsirolimus, Axitinib, and Camptothecin was found relatively higher in the high-risk group, while that of Paclitaxel, Lapatinib, Dasatinib, Bleomycin, Docetaxel, Doxorubicin, Bexarotene, Gefitinib, Bosutinib, and Bortezomib was found relatively higher in the low-risk group. Both results came with statistical significance after log-rank testing (P-value < 0.05).

Taking it all together, it is believed that CRGI not only predicted the OS rate but also implicated those different therapeutic strategies that should be applied for more precise treatment. Therefore, we intended to employ CRGI as a classifier in molecular subtypes analytics. Consequently, it was found that different immunotherapy and chemotherapy had different effects on different molecular subtypes.

In summary, in the present study, we highlighted the outstanding achievement of CRGI in PAAD prognostic prediction and its association with tumor immunology. These findings may benefit future immunotherapy- and chemotherapy-based interventions.

Declarations

Data Availability Statement

The original contributions presented in the study are included in the article and supplementary materials, further resealable inquiries can be directed to the 1st author.

Author Contributions

Conception and design: Xufeng Huang and Shujing Zhou.

Collection and assembly of data: Xufeng Huang and Ying Liu.

Data analysis and visualization: Shujing Zhou, Xufeng Huang, Hajdu András, and János Tóth.

Data interpretation: Xufeng Huang, Shujing Zhou, and Ying Liu.

Manuscript writing: Xufeng Huang and Shujing Zhou, and Ying Liu.

Manuscript revision: Hajdu András, and János Tóth.

Final approval of the manuscript: all authors.

Funding

This work has been supported in part by Project NO.TKP2021-NKTA-34 which has been implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, and financed under the TKP2021-NKTA funding scheme.

Supplementary materials

The supplementary materials for this article can be found online at:

https://drive.google.com/drive/folders/1dIrvCp_WPdQd4mG1g172SdgV5u6hlMMS?usp=sharing

T1_Rare data for the web-based OS calculator

S1_Correlation between the risk score and clinical factors

S2_Correlation with MSI

S3_Comparison of the efficacy of 32 anti-cancerous drugs in high- and low-risk groups

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.

References

1. Tsvetkov, P.; Coy, S.; Petrova, B.; Dreishpoon, M.; Verma, A.; Abdusamad, M.; Rossen, J.; Joesch-Cohen, L.; Humeidi, R.; Spangler, R.D.; et al. Copper induces cell death by targeting lipoylated TCA cycle proteins. Science 2022, 375, 1254-1261.

2. Levy, J., Towers, C. & Thorburn, A. Targeting autophagy in cancer. Nat Rev Cancer 17, 528–542 (2017). https://doi.org/10.1038/nrc.2017.53

3. Lei, G., Zhuang, L. & Gan, B. Targeting ferroptosis as a vulnerability in cancer. Nat Rev Cancer (2022). https://doi.org/10.1038/s41568-022-00459-0

4. Yang, F., Bettadapura, S.N., Smeltzer, M.S. et al. Pyroptosis and pyroptosis-inducing cancer drugs. Acta Pharmacol Sin (2022). https://doi.org/10.1038/s41401-022-00887-6

5. Rossi Sebastiano M, Pozzato C, Saliakoura M, Yang Z, Peng RW, Galie M, et al. Acsl3-Pai-1 Signaling Axis Mediates Tumor-Stroma Cross-Talk Promoting Pancreatic Cancer Progression. Sci Adv (2020) 6(44):1–15. DOI: 10.1126/sciadv.abb9200

6. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J. Clin. 71, 209–249 (2021).

7. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. CA Cancer J. Clin. 71, 7–33 (2021).

8. Miller, K. D. et al. Cancer statistics for adolescents and young adults, 2020. CA Cancer J. Clin. 70, 443–459 (2020).

9. Rawla, P., Sunkara, T. & Gaduputi, V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J. Oncol. 10, 10–27 (2019).

10. Qian, Y. et al. Molecular alterations and targeted therapy in pancreatic ductal adenocarcinoma. J. Hematol. Oncol. 13, 130 (2020).

11. Mizrahi, J. D., Surana, R., Valle, J. W. & Shroff, R. T. Pancreatic cancer. Lancet 395, 2008–2020 (2020).

12. Christenson, E. S., Jaffee, E. & Azad, N. S. Current and emerging therapies for patients with advanced pancreatic ductal adenocarcinoma: a bright future. Lancet Oncol. 21, e135–e145 (2020).

13. Connor, A. A. et al. Integration of genomic and transcriptional features in pancreatic cancer reveals increased cell cycle progression in metastases. Cancer Cell 35, 267–282.e267 (2019).

14. Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., et al. (2019). The International Cancer Genome Consortium Data Portal. Nat. Biotechnol. 37 (4), 367–369. doi:10.1038/s41587-019-0055-9

15. Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL https://www.jstatsoft.org/v33/i01/.

16. Noah Simon, Jerome Friedman, Trevor Hastie, Rob Tibshirani (2011). Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. Journal of Statistical Software, 39(5), 1-13. URL https://www.jstatsoft.org/v39/i05/.

17. Wu, X. et al., 2008. Top 10 algorithms in data mining. Knowledge and information systems, 14(1), pp.1–37.

18. Reynolds D. (2009) Gaussian Mixture Models. In: Li S.Z., Jain A. (eds) Encyclopedia of Biometrics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73003-5_196

19. Wang, Zhe, Rundong Shi, and Shijie Li. “GBDT and BERT: a Hybrid Solution for Recognizing Citation Intent.” (2020).

20. Mucherino, A., Papajorgji, P.J., Pardalos, P.M. (2009). k-Nearest Neighbor Classification. In: Data Mining in Agriculture. Springer Optimization and Its Applications, vol 34. Springer, New York, NY. https://doi.org/10.1007/978-0-387-88615-2_4

21. Ke, G. et al., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, pp.3146–3154.

22. Liaw A, Wiener M (2002). “Classification and Regression by randomForest.” R News, 2(3), 18-22. https://CRAN.R-project.org/doc/Rnews/.

23. Chen, T. & Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. New York, NY, USA: ACM, pp. 785–794. Available at: http://doi.acm.org/10.1145/2939672.2939785.

24. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565–74.

25. ForestPlots.net et al. (2021). Taking the pulse of Earth's tropical forests using networks of highly distributed plots. Biological Conservation 260 (2021) 108849. https://doi.org/10.1016/j.biocon.2020.108849

26. Lopez-Gonzalez, G., Lewis, S.L., Burkitt, M., Baker T.R. and Phillips, O.L. 2009. ForestPlots.net Database. www.forestplots.net.

27. Xu, Q., Wang, Y., Fang, Y. et al. An easy-to-operate web-based calculator for predicting the progression of chronic kidney disease. J Transl Med 19, 288 (2021). https://doi.org/10.1186/s12967-021-02942-y

28. Mootha, V., Lindgren, C., Eriksson, KF. et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34, 267–273 (2003). https://doi.org/10.1038/ng1180

29. Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).

30. Kanehisa, M; Toward understanding the origin and evolution of cellular organisms. Protein Sci. 28, 1947-1951 (2019).

31. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M.; KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 49, D545-D551 (2021).

32. Ashburner et al. Gene ontology: a tool for the unification of biology. Nat Genet. May 2000;25(1):25-9.

33. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. Jan 2021;49(D1):D325-D334.

34. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. Nat. Commun. 4, 2612. doi:10.1038/ncomms3612

35. Taiwen Li#, Jingxin Fu#, Zexian Zeng, David Cohen, Jing Li, Qianming Chen, Bo Li, and X. Shirley Liu. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Research 2020

36. Taiwen Li, Jingyu Fan, Binbin Wang, Nicole Traugh, Qianming Chen, Jun S. Liu, Bo Li, X. Shirley Liu. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Research. 2017;77(21):e108-e110.

37. Bo Li, Eric Severson, Jean-Christophe Pignon, Haoquan Zhao, Taiwen Li, Jesse Novak, Peng Jiang, Hui Shen, Jon C. Aster, Scott Rodig, Sabina Signoretti, Jun S. Liu, X. Shirley Liu. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biology. 2016;17(1):174.

38. Wilkerson, D. M, Hayes, Neil D (2010). “ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.” Bioinformatics, 26(12), 1572-1573. http://bioinformatics.oxfordjournals.org/content/26/12/1572.abstract.

39. Geeleher P, Cox N, Huang RS (2014) pRRophetic: An R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. PLoS ONE 9(9): e107468. https://doi.org/10.1371/journal.pone.0107468

40. Wang, S., Zheng, Y., Yang, F. et al. The molecular biology of pancreatic adenocarcinoma: translational challenges and clinical perspectives. Sig Transduct Target Ther 6, 249 (2021). https://doi.org/10.1038/s41392-021-00659-4

41. Deng, J., Zhang, Q., Lv, L. et al. Identification of an autophagy-related gene signature for predicting prognosis and immune activity in pancreatic adenocarcinoma. Sci Rep 12, 7006 (2022). https://doi.org/10.1038/s41598-022-11050-w

42. Yang, J., Wei, X., Hu, F. et al. Development and validation of a novel 3-gene prognostic model for pancreatic adenocarcinoma based on ferroptosis-related genes. Cancer Cell Int 22, 21 (2022). https://doi.org/10.1186/s12935-021-02431-8

43. Bai Z, Xu F, Feng X, et al. Pyroptosis regulators exert crucial functions in prognosis, progression, and immune microenvironment of pancreatic adenocarcinoma: a bioinformatic and in vitro research. Bioengineered. 2022;13(1):1717-1735. doi:10.1080/21655979.2021.2019873

44. Karasaki, Takahiro, et al. “An Immunogram for the Cancer-Immunity Cycle: Towards Personalized Immunotherapy of Lung Cancer.” Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer vol. 12,5 (2017): 791-803. doi:10.1016/j.jtho.2017.01.005

45. Bukowski K, Kciuk M, Kontek R. Mechanisms of Multidrug Resistance in Cancer Chemotherapy. Int J Mol Sci. 2020;21(9):3233. Published 2020 May 2. doi:10.3390/ijms21093233

46. Goh, Wen Quan Jonathan, et al. “DLAT subunit of the pyruvate dehydrogenase complex is upregulated in gastric cancer-implications in cancer therapy.” American journal of translational research vol. 7,6 1140-51. 15 Jun. 2015

Figures

Figure 1

Graphical abstract of the present study.

Figure 2

Predictive performance comparison of 10 mainstream machine learning algorithms. (A) Distribution of the risk score and survival status in high- and low-risk groups, and the heatmap that depicts the expression of CRGI genes between high- and low-risk groups. (B) Survival analysis of high- and low-risk groups. (C) The time-dependent ROC curve with the AUC value of 1-, 2-, 3-, 4-, and 5-year OS prediction of the best model optimized by LASSO penalized Cox regression. (D)-(L) The predictive accuracy of other machine learning algorithms (i.e., Decision Tree, Gaussian Mixture Model, GBDT, K-NN, LGBM, Logistic Regression, Random Forest, SVM, and XGboost, respectively).

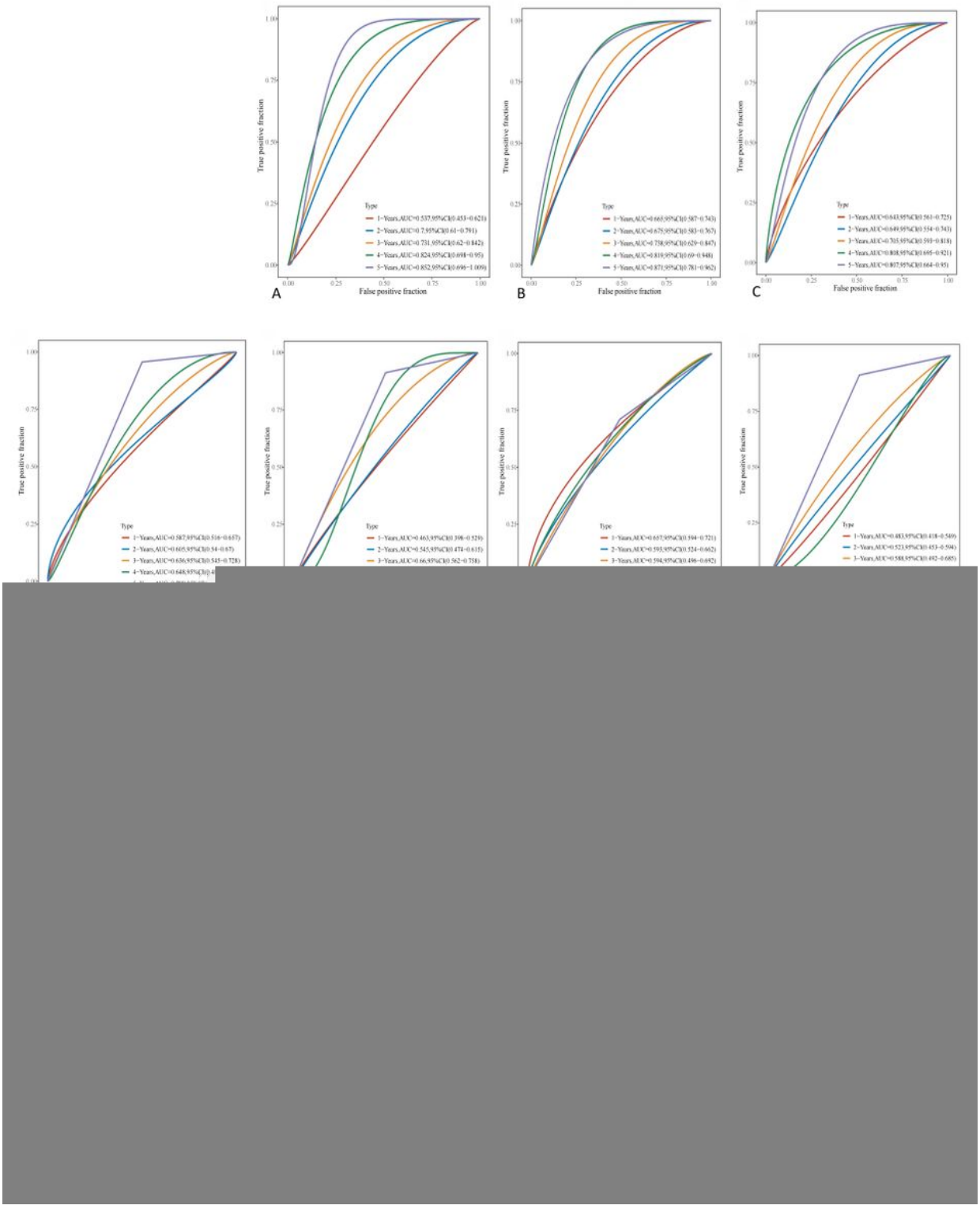

Figure 3

Comparison of other cell death mechanisms-based prognostic signatures in PAAD. (A)-(C) The ROC curve of autophagy-, ferroptosis-, and pyroptosis-based models in the TCGA dataset, respectively. (D)-(G) The ROC curve of our model, and the autophagy-, ferroptosis-, and pyroptosis-based ones in the ICGC dataset, respectively. (H) DCA of our model, and the autophagy-, ferroptosis-, and pyroptosis-based ones in the

TCGA dataset, respectively. (I) DCA of our model, and the autophagy-, ferroptosis-, and pyroptosis-based ones in the ICGC dataset, respectively.

Figure 4

Nomograms with clinicopathological characteristics to predict OS in PAAD. (A)-(B) Forest plots of the results of the univariate and multivariate Cox regression, respectively. (C)-(D) Conventional nomogram and its calibration curve based on our model. (E) The screenshot of the web-based OS calculator that was developed from the underlying statistics and calculating a fictional case.

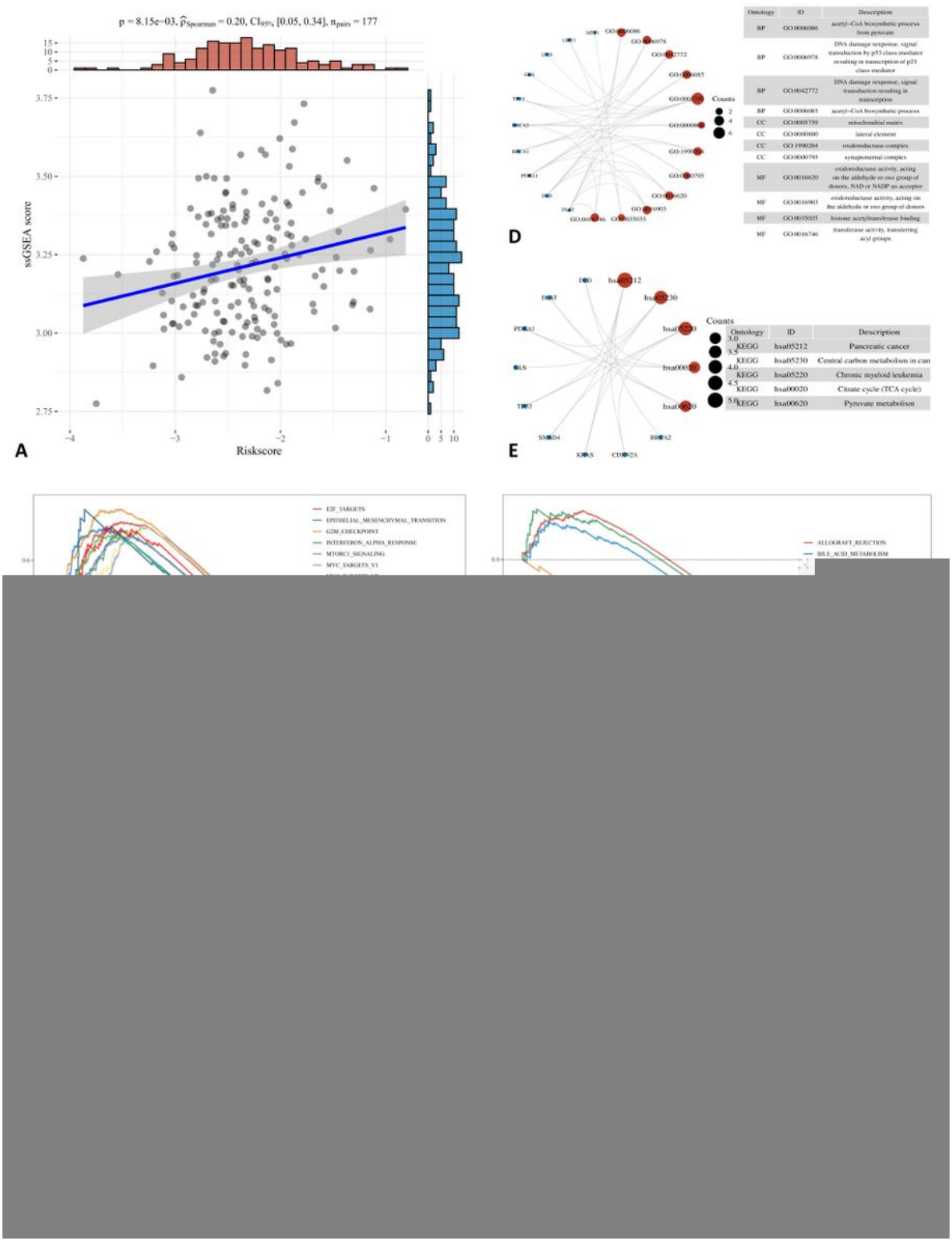

Figure 5

Over-representative analysis of CRGI in PAAD. (A) Correlation analysis of the risk score and ssGSEA score. (B) The top 10 most enriched and upregulated cancerous hallmarks in the high-risk group. (C) The 4 downregulated cancerous hallmarks in the low-risk group. (D) The enriched GO terms of CRGI genes. (E) The enriched KEGG pathways of CRGI genes.

Figure 6

In-depth analytics on the relationship between CRGI and the tumor microenvironment condition, immune cell infiltration, immunotherapy efficacy, TMB, as well as the cancer immunity cycle. (A)-(C) Comparison of the stromal, immune, and ESTIMATE scores of tumorous and normal tissues in the TCGA dataset. (D)-(F) Correlation analysis between the risk score and the stromal, immune, and ESTIMATE scores. (G) The violent plot demonstrated the comparison of the stromal, immune, and ESTIMATE scores in high- and low-risk groups. (H) The heatmap demonstrated the diverse immune cell types in the infiltration process. (I) The violent plot demonstrated the comparison of the expression of the representative genes of the soundest immune checkpoints in high- and low-risk groups. (J) Correlation analysis between the risk score and TIDE score. (K) The violent plot demonstrated the comparison of the TIDE scores in high- and low-risk groups. (L) Correlation analysis between the risk score and TMB score. (M) Graphical demonstration of the cancer immunity cycle. (N) Correlation analysis between the risk score and the main steps of the cancer immunity cycle as well as known biological processes. (O) The boxplot demonstrated the comparison of marker scores in high- and low-risk groups.

Figure 7

Comparison of the efficacy of 32 chemotherapeutics in (A) high-risk group, and (B) low-risk group.

Figure 8

Characterization of 4 CRGI-based molecular subtypes with different survival outcomes and immunotherapy efficacy. (A) The consensus matrix (K= 5), where the columns and rows represented the TCGA samples involved which were clustered into squares in white to deep blue colors. (B) The consensus CDF plot (K = 1, 2, 3, 4, 5, 6). (C) The Delta area plot (K = 1, 2, 3, 4, 5, 6). (D) The PCA diagram where TCGA samples were represented in the manner of scattering spots. The circle bordered the area within which all the samples belonged to the same molecular subtype. (E) The boxplot demonstrated the expression profiles of the 13 CRGI genes in each molecular subtype. (F) Survival analysis demonstrated the clinical outcomes of different molecular subtypes. (G) The boxplot demonstrated the comparison of the TIDE scores in different molecular subtypes. (H) The boxplot demonstrated the expression profiles of the representative genes of the soundest immune checkpoints in different molecular subtypes.

Figure 9

Comparison of the efficacy of 32 chemotherapeutics in 4 CRGI-based molecular subtypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- T1.txt
- S1Correlationbetweentheriskscoreandclinicalfactors.pdf
- S2CorrelationwithMSI.pdf
- S3Comparisonoftheefficacyof32anticancerousdrugsinhighandlowriskgroups.pdf