



OPEN ACCESS

EDITED BY
Chang Zou,
Jinan University, China

REVIEWED BY
Shunwang Cao,
The Second Affiliated Hospital of
Guangzhou University of Chinese
Medicine, China
Xujun He,
Zhejiang Provincial People's Hospital,
China

*CORRESPONDENCE
Zhengping Yu,
yuzhengping@wzhospital.cn

[†]These authors have contributed equally
to this work

SPECIALTY SECTION
This article was submitted to Signaling,
a section of the journal
Frontiers in Cell and Developmental
Biology

RECEIVED 16 August 2022
ACCEPTED 30 August 2022
PUBLISHED 19 September 2022

CITATION
Zhu K, Tao Q, Yan J, Lang Z, Li X, Li Y,
Fan C and Yu Z (2022), Machine learning
identifies exosome features related to
hepatocellular carcinoma.
Front. Cell Dev. Biol. 10:1020415.
doi: 10.3389/fcell.2022.1020415

COPYRIGHT
© 2022 Zhu, Tao, Yan, Lang, Li, Li, Fan
and Yu. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Machine learning identifies exosome features related to hepatocellular carcinoma

Kai Zhu^{1†}, Qiqi Tao^{1†}, Jiatao Yan^{2†}, Zhichao Lang¹, Xinmiao Li¹, Yifei Li¹, Congcong Fan¹ and Zhengping Yu^{1*}

¹Department of Hepatobiliary Surgery, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, ²Wenzhou Business College, Wenzhou, China

Background: Hepatocellular carcinoma (HCC) is one of the most malignant tumors with a poor prognosis. There is still a lack of effective biomarkers to predict its prognosis. Exosomes participate in intercellular communication and play an important role in the development and progression of cancers.

Methods: In this study, two machine learning methods (univariate feature selection and random forest (RF) algorithm) were used to select 13 exosome-related genes (ERGs) and construct an ERG signature. Based on the ERG signature score and ERG signature-related pathway score, a novel RF signature was generated. The expression of BSG and SFN, members of 13 ERGs, was examined using real-time quantitative polymerase chain reaction and immunohistochemistry. Finally, the effects of the inhibition of BSG and SFN on cell proliferation were examined using the cell counting kit-8 (CCK-8) assays.

Results: The ERG signature had a good predictive performance, and the ERG score was determined as an independent predictor of HCC overall survival. Our RF signature showed an excellent prognostic ability with the area under the curve (AUC) of 0.845 at 1 year, 0.811 at 2 years, and 0.801 at 3 years in TCGA, which was better than the ERG signature. Notably, the RF signature had a good performance in the prediction of HCC prognosis in patients with the high exosome score and high NK score. Enhanced BSG and SFN levels were found in HCC tissues compared with adjacent normal tissues. The inhibition of BSG and SFN suppressed cell proliferation in Huh7 cells.

Conclusion: The RF signature can accurately predict prognosis of HCC patients and has potential clinical value.

KEYWORDS

TCGA database, exosome, PI3K-Akt pathway, immune cell, machine learning, prognosis signature, HCC

Introduction

Hepatocellular carcinoma (HCC) is the major histologic subtype of liver cancer, accounting for 90% of primary liver cancer. Currently, HCC is the third leading cause of cancer-related death worldwide, with the increasing incidence (Sung et al., 2021). HCC is mainly caused by viral infection, carcinogen, and chronic inflammation (Schulze et al., 2016). Despite improvements in systemic therapy, surgical technique, and radiation therapy, the overall 5-year survival rate of HCC patients is currently less than 20% (Fujiwara et al., 2018). Therefore, identifying prognostic markers for HCC is important to prolong patient survival time.

Exosomes are small extracellular nanovesicles with diameters of 30–150 nm that play an important role in cellular communication and epigenetic regulation by transporting key proteins and genetic material (e.g., miRNA, mRNA, and DNA) (Yang et al., 2020). Cancer cell-secreted exosomes, which are involved in cancer cell genesis and development, may promote tumor proliferation and metastasis by participating in cellular communication, regulating cell signaling, and promoting the formation of a pre-metastatic niche (Kharaziha et al., 2012; Hoshino et al., 2015; Peinado et al., 2017; Wortzel et al., 2019; Zhang and Yu, 2019). It has been shown that exosomes from cancer cells regulate antitumor immune response by inhibiting T-cell activation and proliferation, inducing regulatory T cells and bone marrow-derived suppressor cells, and suppressing natural killer (NK) and CD8⁺ T-cell functions (Wieckowski et al., 2009; Pyzer et al., 2016; Whiteside, 2016). Exosomal protein markers as well as nucleic acid markers can be used for early diagnosis and assessment of the prognostic risk of recurrence in a variety of cancers (Mashouri et al., 2019; Li et al., 2021). For example, the levels of exosomal CD151, CD171, and TSPAN8 have been proven to be effective indicators for predicting the prognosis of lung cancer (Iqbal et al., 2019). Nucleic acids such as miR-21 and miR-1246 in exosomes can be used as markers for breast cancer detection (Dong et al., 2020). Although the predictive significance of exosomes in HCC has been investigated, there are no reports on the systematic detection of exosomal biomarkers in patients with HCC.

In this study, we used two machine learning methods (univariate feature selection and random forest (RF) algorithm) to screen out 13 exosome-related genes (ERGs) and constructed an ERG signature with significant prognostic prediction performance. Then, enrichment analysis identified some remarkable factors that correlated with the prognosis of HCC patients. Patients with both high-risk exosome scores and high-risk NK scores had the worst survival. Finally, ERG, mitosis, PI3K-Akt pathway, B cell, NK cell, and CD8⁺ T cell were integrated into an RF signature by an RF algorithm. This RF signature was shown to have a better prognostic prediction performance than any single factor. Then, real-time quantitative polymerase chain reaction (PCR) and immunohistochemistry (IHC) were performed to verify the differential expressions of prognostic genes in HCC tissues and adjacent normal tissues.

Materials and methods

Data collection

The patient data from TCGA LIHC (<https://portal.gdc.cancer.gov/projects/TCGA-LIHC>) included 424 cases. The samples included 374 tumor samples and 50 normal samples. Samples from the International Cancer Genome Consortium (ICGC) were downloaded at <https://dcc.icgc.org/releases/current/Projects/LIRI-JP>. Among them, RNA-seq data were available for 231 cases, of which 9 patients presented multiple duplicates of tumor samples. ERGs were obtained at <http://www.exocarta.org/>, with a total of 2700 genes (Supplementary Table S1). The obtaining of 30 pairs of HCC and adjacent tumor tissue samples from The First Affiliated Hospital of Wenzhou Medical University (FAHWMU) was approved by the FAHWMU Ethics Committee. Moreover, written informed consent was also obtained from the participating patients.

Gene selection and construction of the ERG signature

We first used univariate feature selection to identify the best genes correlated with the overall survival (OS) time using Python 3.9.7. Univariate feature selection is a simple method for evaluating the importance of features using univariate statistical tests. Then, we built a classifier using RF algorithm, and genes with feature importance >0.06 were selected (Breiman, 2004). RF is a classifier that contains multiple decision trees. Its output category is determined by the mode of the category of the output of the individual tree. Univariate Cox regression analysis was used to identify whether those selected genes are independent prognostic factors ($p < 0.05$). Least absolute shrinkage and selection operator (LASSO) was then used for further selection. The optimum penalty parameter (λ) for the model was determined by ten-fold cross-validation following the minimum criteria (i.e., the value of λ corresponding to the lowest partial likelihood deviance). The correlation coefficient was calculated using the Spearman correlation analysis. The finally selected genes in TCGA cohort included: LDHA, PRDX1, HADHA, SFN, UBB, S100A10, BSG, LRP4, HSPA8, CALM1, PON1, IDH1, and PRDX6.

ERG scores of the patients were calculated according to the normalized expression of each gene and its corresponding coefficient from the feature importance calculated by RF algorithm. The formula was established as follows:

$$\text{ERG score} = e^{\sum (\text{expression level of each gene} \times \text{corresponding coefficient})}$$

Functional enrichment

Patients were stratified into high- or low-risk groups based on the median value of their ERG score. Then, we screened different expression genes (DEGs) between low- and high-risk

groups by the “limma” package of R ($|\log_2FC| \geq 1$, FDR < 0.05). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses for DEGs were performed using the OmicShare tools (<https://www.omicshare.com/tools>). The protein–protein interaction (PPI) networks of DEGs were generated by the STRING database (STRING: [functional protein association networks \(string-db.org\)](https://string-db.org)).

The “prcomp” function of the “stats” R package was used to perform principal component analysis (PCA). Additionally, using the “Rtsne” R package, we explored the clustering of different groups using t-distributed stochastic neighbor embedding (t-SNE). In order to determine the predictive accuracy of the model, the “pROC” R package was used to analyze receiver operating characteristic (ROC) curves. The survival analysis was performed by the Kaplan–Meier method. Using CIBERSOFT algorithm, the relative content score of 22 tumor immune-infiltrated cells in each HCC patient was calculated. The heatmap was drawn using the “pheatmap” package of R.

Gene normalization

The genes used for calculating the mitosis score, NK score, PI3K-Akt score, B cell score, and CD8⁺ T score were downloaded from GSEA (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>) or previous literature works (Bolotin et al., 2017; Hydring et al., 2017; Mariathasan et al., 2018; Dyugay et al., 2022) (Supplementary Table S1). All those genes were normalized by LASSO and univariate Cox regression to improve the accuracy of prognosis prediction (Supplementary Table S2).

Construction of the machine learning signature

Random forest modeling is one of the most universal machine-learning algorithms, which can model response prediction via fitting training data on the basis of different input features. It was performed using the Random Forest Classifier from the Python scikit-learn library. The hyperparameters of the model (i.e., the maximal amount of samples in the leaf and tree depth) were optimized with the use of RandomizedSearchCV and GridSearchCV from the Python scikit-learn library with 5-fold cross-validation. The F1 score was used as a measure of quality to train the model:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

The variance of the RF model can be explained as the difference between training and test set quality metrics.

To explore input feature importance and interactions, the RF model was applied to all patients with Shapley additive explanation (SHAP) (Lundberg et al., 2020). SHAP is a game theory method that interprets the machine-learning model and understands the

decision-making process by quantifying the contribution that each feature brings to the prediction made by the model.

Quantitative real-time PCR analysis

We performed different analyses for the 13 ERGs between HCC samples and normal liver samples (Supplementary Table S3). BSG, S100A10, and SFN had the smallest FDR value. In addition, BSG and SFN were shown to have a significant prognosis for HCC patients (Supplementary Figure S1). In this case, we chose BGS and SFN for further analysis.

We collected 15 pairs of HCC and adjacent tumor tissue samples from the FAHWMU. Total RNA was isolated from tumor tissues as well as adjacent normal tissues using the Tiangen RNA extraction reagent kit. Each sample was reversely transcribed into complementary DNA (cDNA) using a reverse-transcription (RT) reagent kit (Takara Biotechnology Co., Ltd., Dalian, China). Then, real-time PCR was performed using SYBR Premix ExTaq (Takara). GAPDH was used as an endogenous control for mRNAs.

Immunohistochemistry

We collected 15 pairs of HCC and adjacent tumor tissue samples from the FAHWMU. The tissues were immersed in 10% formalin for fixation, and then the formalin-fixed tissues were dehydrated and rehydrated. Next, the sections, blocked in 10% BSA, were incubated with anti-SFN and anti-BSG primary antibodies at 4°C for at least 12 h. Then, the sections were incubated with a horseradish peroxidase-conjugated secondary antibody for 30 min. Finally, quantitative analysis for SFN- and BSG-positive areas was performed under the microscope (Carl Zeiss, Germany).

Cell culture

The human HCC cell line Huh7 was purchased from ATCC. Huh7 was cultured in DMEM with 10% fetal bovine serum (FBS) and 1% antibiotics. The cells were maintained in a 37°C incubator with 5% CO₂.

Cell transfection

The Huh7 cell line was cultured in a six-well plate at a density of 8×10^3 cells per well. When the cell density was near to 50%, si-NC, si-SFN, and si-BSG packaged by lipo2000 were transfected into cells at 37°C for 6 h. Then, the fresh medium was replaced, and the cells were collected for subsequent experiments after 48 h of transfection.

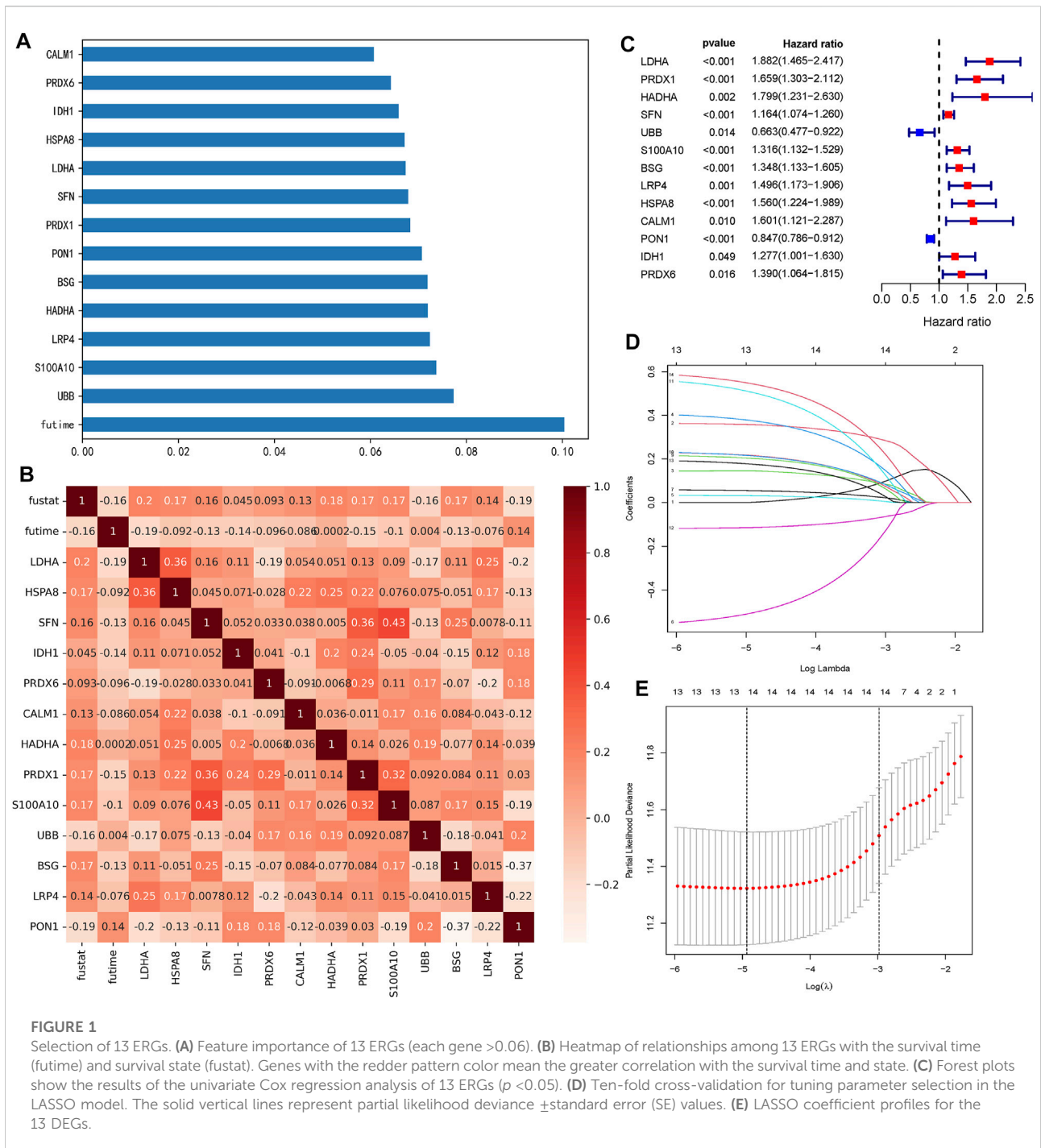


FIGURE 1

Selection of 13 ERGs. (A) Feature importance of 13 ERGs (each gene >0.06). (B) Heatmap of relationships among 13 ERGs with the survival time (future) and survival state (fustat). Genes with the redder pattern color mean the greater correlation with the survival time and state. (C) Forest plots show the results of the univariate Cox regression analysis of 13 ERGs ($p < 0.05$). (D) Ten-fold cross-validation for tuning parameter selection in the LASSO model. The solid vertical lines represent partial likelihood deviance \pm standard error (SE) values. (E) LASSO coefficient profiles for the 13 DEGs.

Cell proliferation assay

Cell counting kit-8 (CCK-8) (Dojindo, Japan) was used for the assessment of cell proliferation. Cells were seeded into a 24-well plate at a density of $1 \times 10^5/100$ ml per well and incubated for 48 h. Then, 100 μ l CCK-8 solution was added to each well and maintained in a 37°C incubator for 1 h. Finally, the absorbance of each well was measured at 450 nm.

Statistical analysis

R software (version 4.0.3) and Python software (version 3.9.7) were used to complete all statistical work and plot drawings. Survival plots were created using the Kaplan–Meier estimator. The AUC was calculated with the “pROC” R package. For comparing multiple survival curves, univariate Cox regression analysis was used. The relative reliability of the RF

signature was estimated by the Akaike information criterion (Akaike, 1974) and concordance index (Lin, 1989). Differences between two groups were compared using Student's *t*-test. $p < 0.05$ was considered significant.

Results

Construction and verification of the ERG signature

From 2700 ERGs, we selected 13 genes with feature importance >0.06 (Figure 1A). The Spearman correlation coefficients among 13 ERGs are shown in Figure 1B. Moreover, the relationships between 13 ERGs and OS time as well as the survival status were also analyzed (Figure 1B). Clearly, all these ERGs were significantly correlated with the survival status. The results of the univariate Cox regression analysis indicated that all 13 ERGs were independent prognostic factors for HCC ($p < 0.05$). These ERGs passed LASSO selection ($\lambda_{\min} = -4.9$ and $\lambda_{1se} = -3$) (Figures 1C,D). The ERG score was calculated using mRNA expression levels and relevant coefficients of these 13 ERGs.

Patients in TCGA and ICGC cohorts were then divided into the high- or low-risk groups according to the median cut-off value, respectively. The results of the Kaplan–Meier curve indicated that patients in the low-risk group exhibited a significantly better OS than those in the high-risk group in TCGA (Figure 2A, $p = 1.342 \times 10^{-13}$) and ICGC cohorts (Figure 2E, $p = 0.0206$). The predictive performance of the ERG score for OS was evaluated by time-dependent ROC curves. In TCGA cohort, the area under the curve (AUC) reached 0.820 at 1 year, 0.791 at 2 years, and 0.786 at 3 years (Figure 2B). In the ICGC cohort, the AUC was 0.662 at 1 year, 0.639 at 2 years, and 0.666 at 3 years (Figure 2D). The PCA and t-SNE plots showed that the ERG score can distinguish the low-risk group from the high-risk group in TCGA cohort (Figures 2C,F). The result of the univariate Cox regression analysis indicated that the ERG score and stage of patients were independent prognostic factors (Figure 2G, $p < 0.001$).

Identification of the mitosis score

DEGs were subsequently identified in the high- and low-risk groups. As presented in Figure 3A, GO analysis showed that the differential genes were mainly enriched in mitotic-related processes (mitotic sister chromatid segregation, mitotic nuclear division, sister chromatid segregation, and cell proliferation, etc.). The associations between these DEGs were analyzed and clustered through the STRING website. The largest cluster was mainly enriched in mitotic-related genes (Figure 3B). Full cluster genes are listed in Supplementary Table S4.

A study has proposed a mitosis-related gene signature for tumors (Hydbring et al., 2017). Genes for the mitosis-related gene signature are shown in Supplementary Table S1. Then, we applied this signature to HCC patients in TCGA cohort. The Kaplan–Meier curve showed that a lower score of the signature was associated with better HCC OS (Figure 3D, $p < 0.05$). The AUC was 0.734 at 1 year, 0.676 at 2 years, and 0.658 at 3 years (Figure 3C). Then, we used univariate Cox regression and LASSO analyses to normalize mitosis-related genes (BRSK1, CDKN2B, GML, KIF2C, RAN, and TTK). The mitosis signature had a better performance in predicting HCC prognosis after normalizing (Figure 3F, $p < 0.05$). Also, the AUC reached 0.750 at 1 year, 0.710 at 2 years, and 0.676 at 3 years (Figure 3E), which was better than the non-normalized signature.

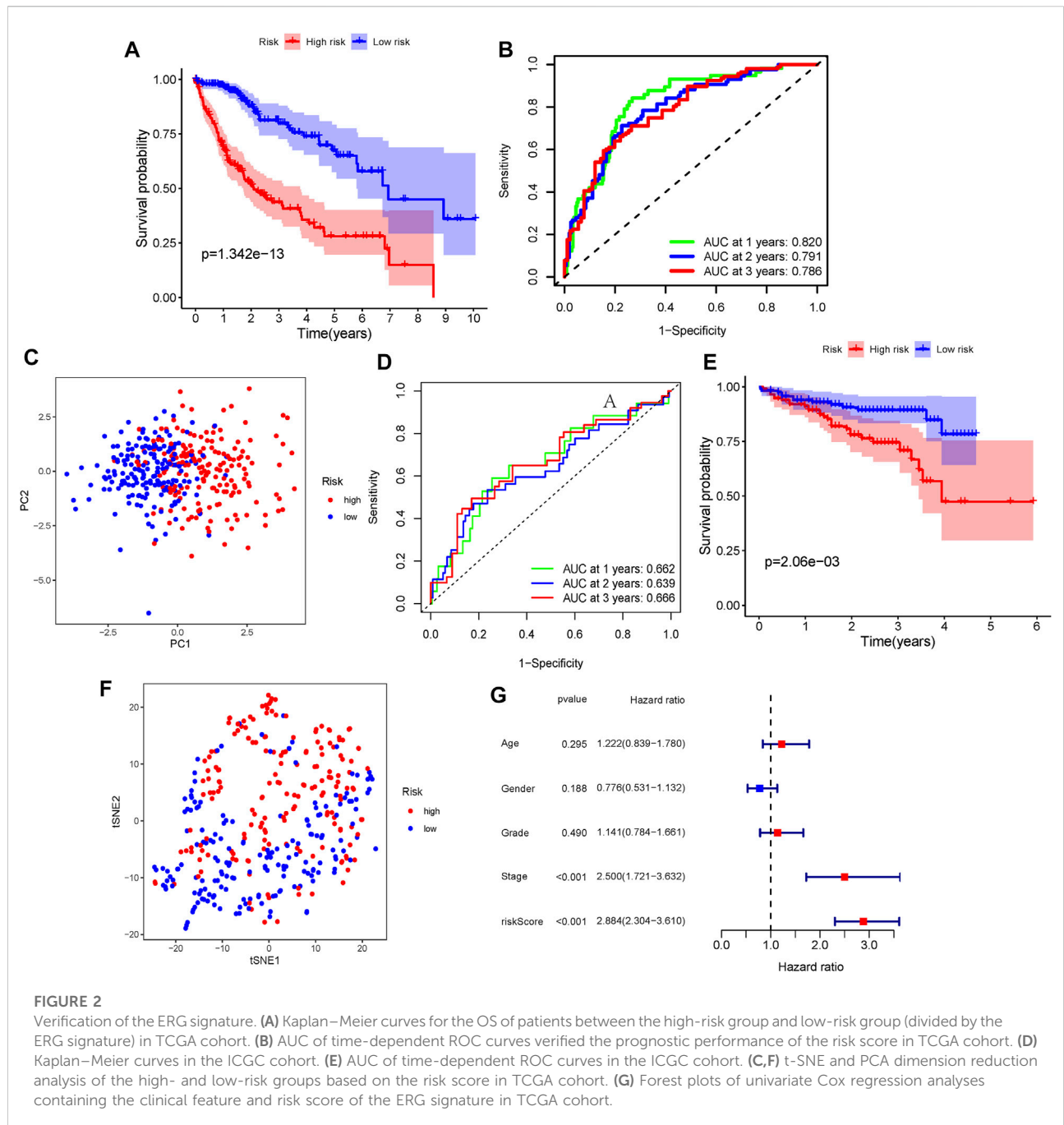
Immune-related processes were selected

The differential genes between low- and high-risk groups were engaged in both intercellular and extracellular processes (Figure 4B), indicating that our ERG signature may play multi-roles in the cell-to-cell interaction process. The results of the KEGG analysis for differential genes suggested that the immune system had the most enrichment of pathway annotation (Figure 4A). Detailed pathway annotation results are shown in Supplementary Table S5. Then, 22 tumor immune-infiltrated cells were calculated for each HCC patient in TCGA cohort (Figure 4C). In addition, our analysis further confirmed that the processes related to PI3K-Akt, B cell, NK cell, and CD8⁺ T cell were identified as the best prognostic relevance in all immune system pathways, which was used for further analysis.

Construction and verification of the RF signature

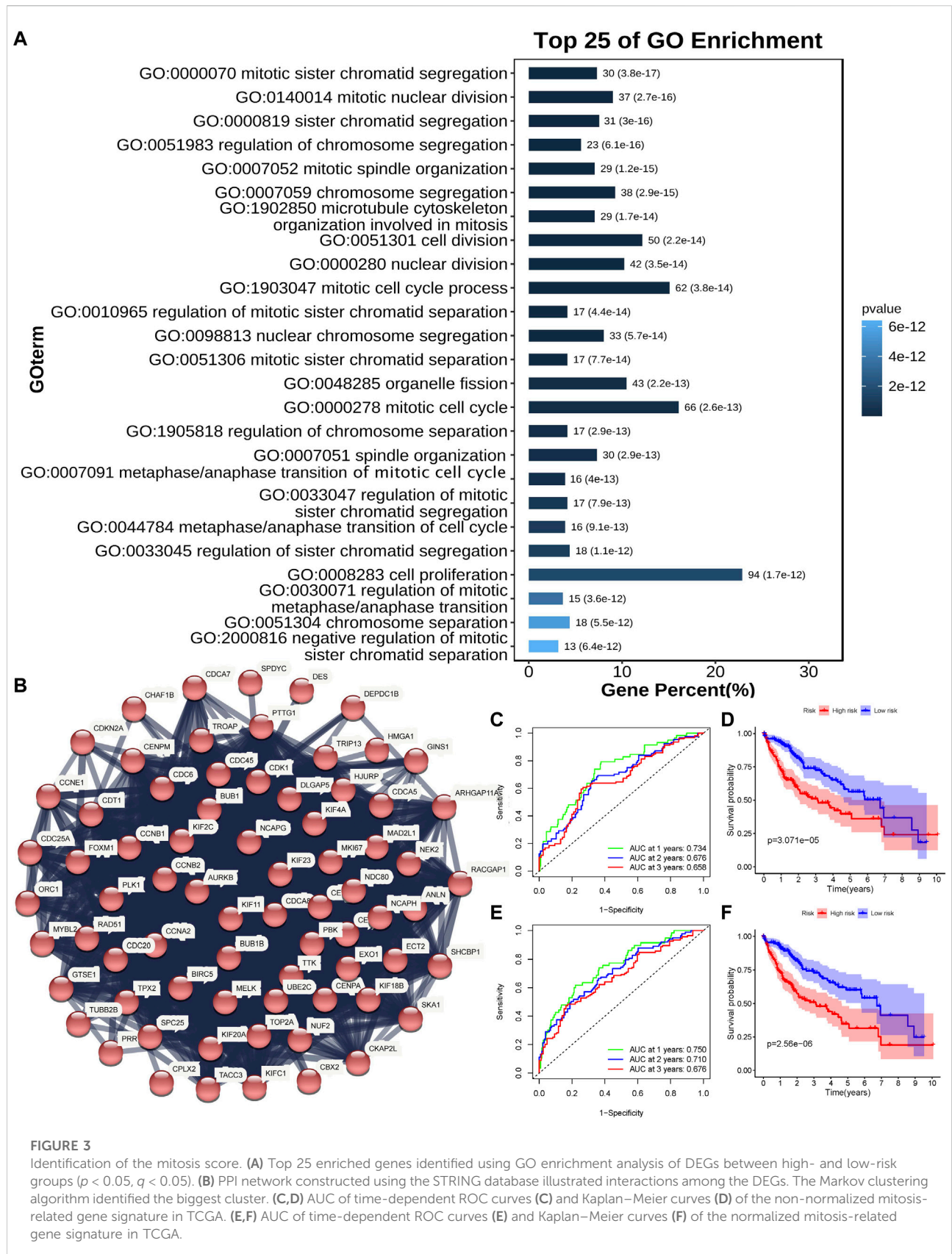
To more accurately predict HCC prognosis, we constructed an RF model via a nested cross-validation approach. We expected features selected for the final iteration of the RF signature to have higher feature importance for response compared with randomly generated numbers. Before selecting the final set of input features, we normalized each immune-related gene set by univariate Cox regression and LASSO analyses. Only the genes that were significantly associated with the patients' prognosis could be included (Supplementary Table S2). Finally, we integrated the ERG score (also named as exosome score), the mitosis score, the PI3K-Akt score, the B-cell score, the NK-cell score, and the CD8⁺ T-cell score to establish a new RF signature.

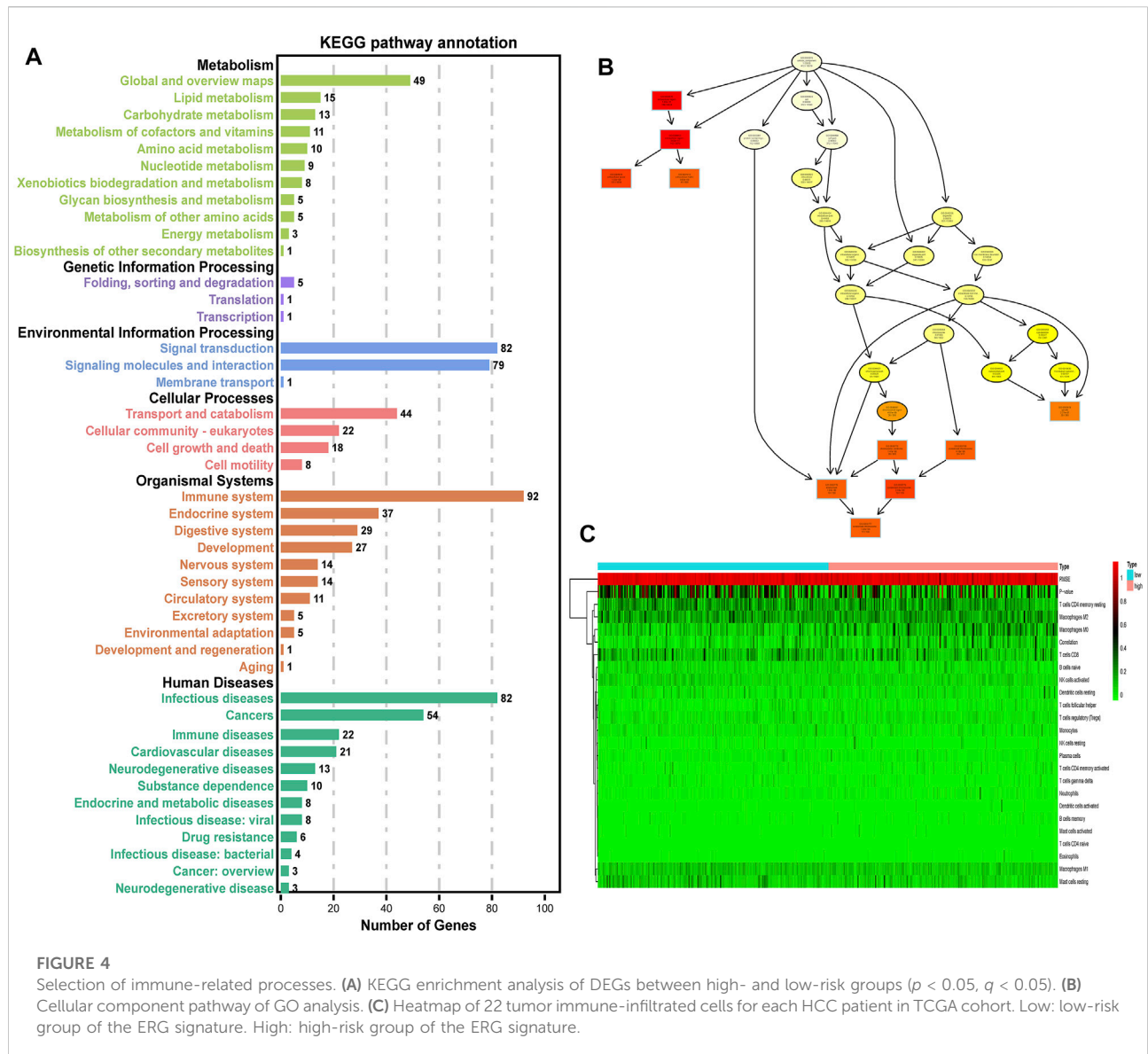
Next, we compared the contribution of our input feature with randomly generated numbers. Notably, each RF feature had higher feature importance than randomly generated numbers, and the exosome score had the most significant association with



response (Figures 5A,B). We also identified the interactions between the variables. We found that the high expression of the exosome score combined with the high expression of NK score had a better predictive performance (Figures 5C,D). These data suggest that our RF signature has the best prognostic accuracy for those HCC patients who have both high exosome scores and high NK scores among all the groups (Figure 6A). Furthermore, the patients with both high-risk exosome scores and high-risk NK scores had the worst survival prognosis.

Moreover, we found an association between the high expression of exosome score and high expression of PI3K-Akt score (Figure 5E), indicating that the RF signature has the best predictive performance for those patients who have both high-risk PI3K-Akt scores and high-risk exosome scores. Furthermore, the patients with both high-risk PI3K-Akt scores and high-risk exosome scores also had the worst survival prognosis among all the groups (Figure 6B). Additionally, the value of the PI3K-Akt score is shown in Figure 5F.





The RF signature showed a better predictive performance than the ERG signature (Figure 6D, $p = 6.55 \times 10^{-15}$), with an AUC of 0.845 for 1 year, 0.811 for 2 years, and 0.801 for 3 years (Figure 6C) in TCGA cohort. In the ICGC cohort, the RF signature also had a better performance (Figure 6F, $p = 4.707 \times 10^{-4}$), with an AUC of 0.733 for 1 year, 0.713 for 2 years, and 0.749 for 3 years (Figure 6E).

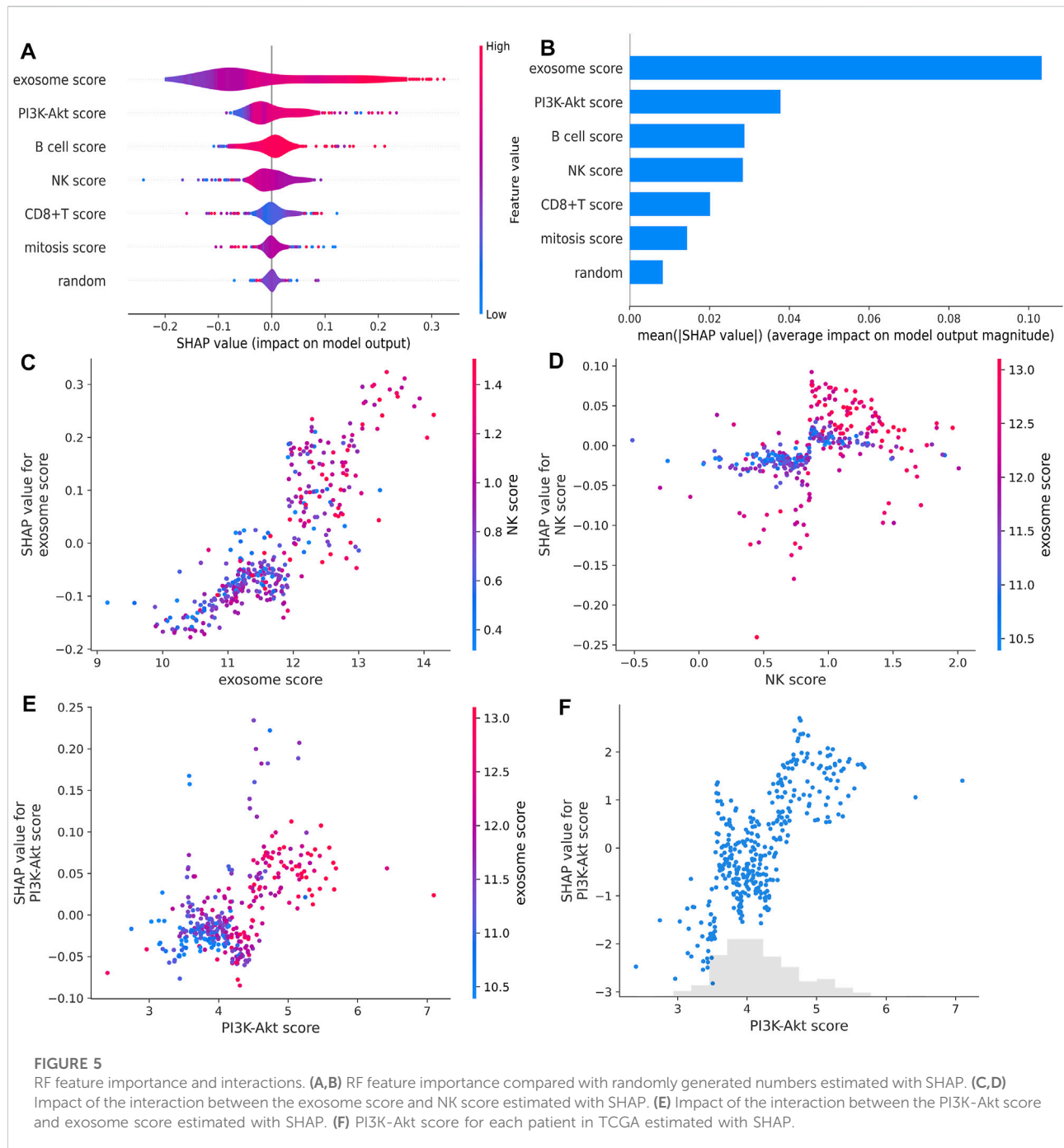
Validation of the expression of prognostic genes by quantitative real-time PCR and IHC

Whether these prognostic genes are dysregulated in HCC was further examined. BSG and SFN, members of 13 ERGs, were

examined in HCC tissues in an independent sample cohort via quantitative real-time PCR and IHC. We found that BSG and SFN were highly expressed in HCC samples compared with adjacent normal tissues (Figures 7A,B). In line with it, similar results were shown in IHC, which were consistent with the mRNA results of BSG and SFN in TCGA and ICGC.

Effects of inhibition of BSG and SFN on cell proliferation

The effects of inhibition of BSG and SFN on cell proliferation were next examined using CCK-8. Our results showed that the knockdown of BSG as well as SFN inhibition led to a significant reduction in cell proliferation (Figure 7C).



Discussion

It has been reported that tumor cells have a higher mitotic frequency than normal cells (Tao, 2005). Mitosis is confirmed as an anti-tumor target. In fact, drugs that stop cells in mitosis, also called anti-mitoses, are common treatments for a variety of human cancers, including breast, non-small-cell lung, and ovarian cancers (Weaver and Cleveland, 2005). Several studies have reported that exosomes can affect the mitosis of tumor

cells. For example, Oliver et al. found that exosomes derived from human primed mesenchymal stem cells induce tumor cell mitosis (Yuan et al., 2019), and similar results could be found in colon cancer cells (Ren et al., 2019). In addition, exosomes produced by tumor cells have been demonstrated to play an important role in immune regulation. Exosomes have been demonstrated to promote inflammation, infectious diseases, and autoimmunity *via* regulating immune stimulation or suppression (Robbins and Morelli, 2014). All

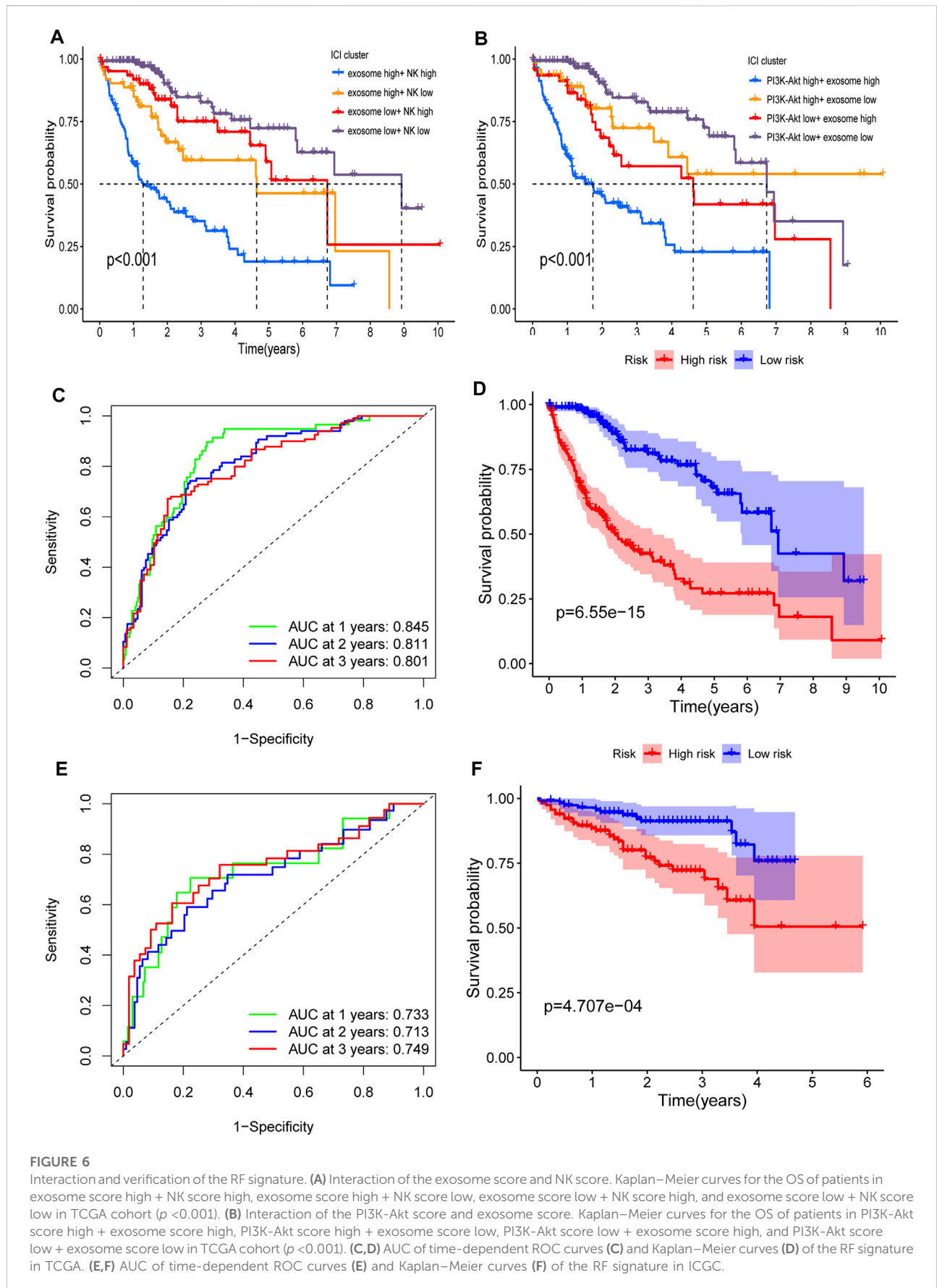


FIGURE 6

Interaction and verification of the RF signature. **(A)** Interaction of the exosome score and NK score. Kaplan–Meier curves for the OS of patients in exosome score high + NK score high, exosome score high + NK score low, exosome score low + NK score high, and exosome score low + NK score low in TCGA cohort ($p < 0.001$). **(B)** Interaction of the PI3K-Akt score and exosome score. Kaplan–Meier curves for the OS of patients in PI3K-Akt score high + exosome score high, PI3K-Akt score high + exosome score low, PI3K-Akt score low + exosome score high, and PI3K-Akt score low + exosome score low in TCGA cohort ($p < 0.001$). **(C,D)** AUC of time-dependent ROC curves **(C)** and Kaplan–Meier curves **(D)** of the RF signature in TCGA. **(E,F)** AUC of time-dependent ROC curves **(E)** and Kaplan–Meier curves **(F)** of the RF signature in ICGC.

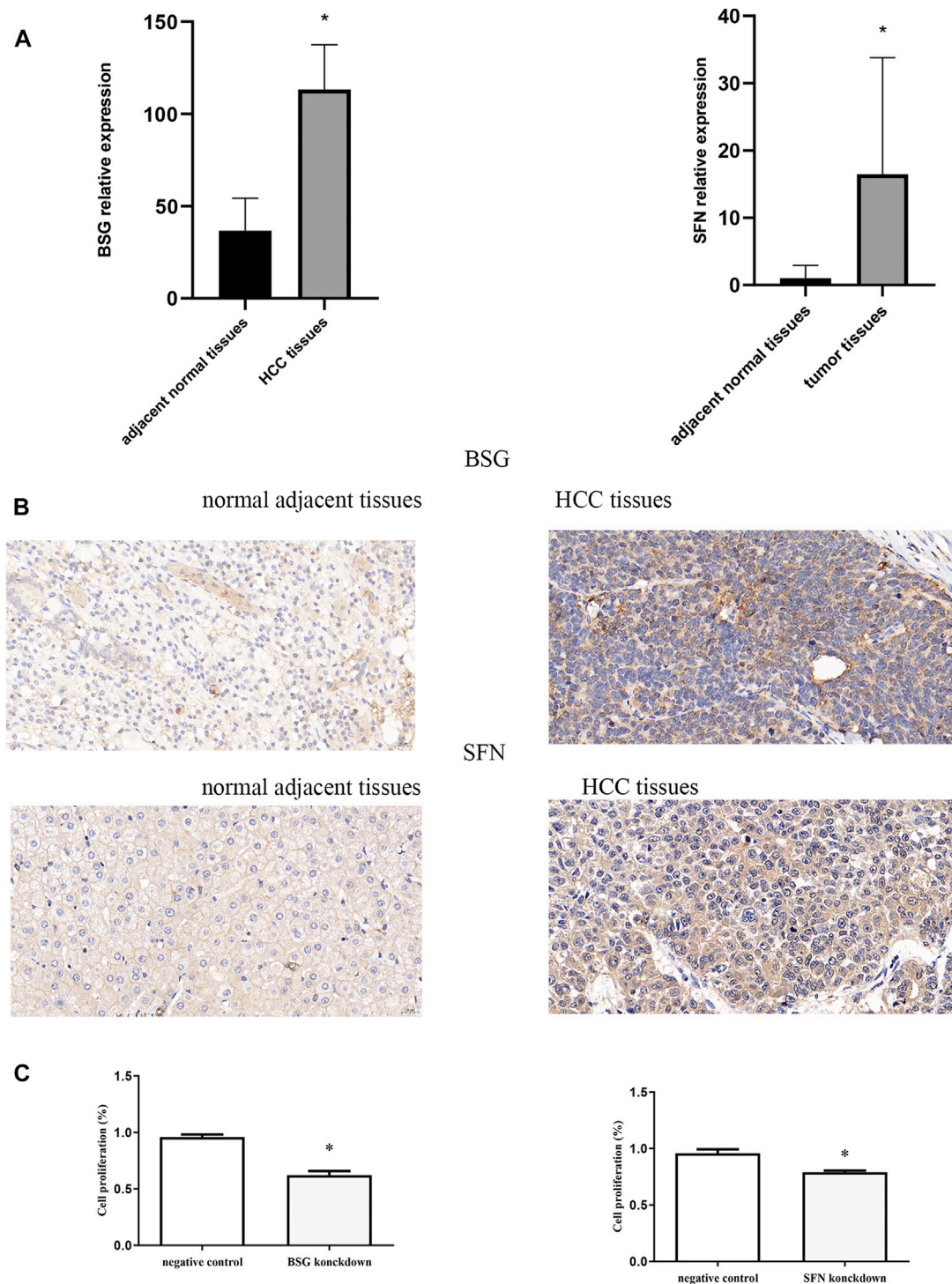


FIGURE 7 Quantitative real-time PCR, immunohistochemistry, and CCK-8. **(A)** mRNA expression of BSG and SFN. **(B)** Levels of BSG and SFN were analyzed by IHC. The scale bar represents 20 μ m. **(C)** Cell proliferation * $p < 0.05$.

aforementioned evidence suggests that a combination of exosome, mitosis, and immunity may have potential predictive values for HCC.

In this study, we used two machine learning methods (univariate feature selection and RF algorithm) to screen out 13 ERGs, which were used to construct an ERG signature. Finally, this ERG signature

was demonstrated to have a significant prognostic prediction performance. Then, enrichment analysis identified some remarkable factors that correlated with the prognosis of HCC patients. We found that patients with high-risk exosome scores and high-risk NK scores had the worst survival. Subsequently, ERG, mitosis, PI3K-Akt pathway, B cell, NK cell, and CD8⁺ T cell were integrated into an RF signature via RF algorithm. This RF signature was demonstrated to have a better prognostic prediction performance than any single factor. In HCC tissues, we also assessed the dysregulation of BSG and SFN, members of key prognostic genes of HCC. Clearly, it was found that BSG and SFN were enhanced in HCC tissues. Inhibition of BSG or SFN led to a reduction in HCC proliferation.

In fact, many prognostic and predictive signatures for tumor patients based on exosome-related processes have been developed (Wu et al., 2021; Li et al., 2022; Lin et al., 2022; Wu et al., 2022). But, taking only exosome-related processes into consideration may not be enough. Recently, it has been reported that immune-related responses are also included in tumor prognostic signatures. Combined with these, an effective predictive model should take different components of the immune system into consideration, and its predictive response may change according to different cancer types (Bruni et al., 2020). Increasing evidence has shown the involvement of B cells (Sautes-Fridman et al., 2019; Cabrita et al., 2020; Helmink et al., 2020; Petitprez et al., 2020; Sharonov et al., 2020), NK cells (Andre et al., 2018), and CD8⁺ T cells (Mariathasan et al., 2018) in cancer immunotherapy responses and immunosurveillance. In our study, we combined B cell, NK cell, and CD8⁺ T cell parameters with the tumor microenvironment, exosome, mitosis, and PI3K-Akt pathway. In addition, to the best of our knowledge, this is the first report to use a novel machine learning method (RF algorithm) to construct this signature.

However, this study still has some disadvantages. For example, our RF signature was only validated in TCGA and ICGC datasets, and more external validation sets should be included. In addition, exosome-, mitosis-, and immune-related processes, which are used to construct the RF model, should have more experimental validation to elaborate the relationships among them.

In conclusion, we constructed an RF signature that contained exosome, mitosis, PI3K-Akt, B cell, NK cell, and CD8⁺ T cell signaling. The RF signature can accurately predict the prognosis of HCC patients and has potential clinical value.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession numbers can be found in the article/[Supplementary Material](#).

Ethics statement

The studies involving human participants were reviewed and approved by the Human Research Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

Author contributions

KZ and QT conceived the project idea and wrote the manuscript. JY, CF, and YL participated in data analysis. ZL, XL, and ZY participated in discussion and language editing. All authors contributed to the manuscript and approved the submitted version.

Funding

The project was supported by the Zhejiang Provincial Research Centre for Diagnosis and Treatment of Critical Liver and Biliary Diseases (Minimally Invasive).

Acknowledgments

The authors thank TCGA and ICGC databases for providing valuable datasets.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2022.1020415/full#supplementary-material>

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19, 716–723. doi:10.1109/tac.1974.1100705
- Andre, P., Denis, C., Soulas, C., Bourbon-Caillet, C., Lopez, J., Arnoux, T., et al. (2018). Anti-NKG2A mAb is a checkpoint inhibitor that promotes anti-tumor immunity by unleashing both T and NK cells. *Cell* 175 (7), 1731–1743 e13. doi:10.1016/j.cell.2018.10.014
- Bolotin, D. A., Poslavsky, S., Davydov, A. N., Frenkel, F. E., Fanchi, L., Zolotareva, O. I., et al. (2017). Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.* 35 (10), 908–911. doi:10.1038/nbt.3979
- Breiman, L. (2004). Random forests. *Mach. Learn.* 45, 5–32.
- Bruni, D., Angell, H. K., and Galon, J. (2020). The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat. Rev. Cancer* 20 (11), 662–680. doi:10.1038/s41586-020-0285-7
- Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., et al. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* 577 (7791), 561–565. doi:10.1038/s41586-019-1914-8
- Dong, X., Bai, X., Ni, J., Zhang, H., Duan, W., Graham, P., et al. (2020). Exosomes and breast cancer drug resistance. *Cell Death Dis.* 11 (11), 987. doi:10.1038/s41419-020-03189-z
- Dyugay, I. A., Lukyanov, D. K., Turchaninova, M. A., Serebrovskaya, E. O., Bryushkova, E. A., Zaretsky, A. R., et al. (2022). Accounting for B-cell behavior and sampling bias predicts anti-PD-L1 response in bladder cancer. *Cancer Immunol. Res.* 10 (3), 343–353. doi:10.1158/2326-6066.cir-21-0489
- Fujiwara, N., Friedman, S. L., Goossens, N., and Hoshida, Y. (2018). Risk factors and prevention of hepatocellular carcinoma in the era of precision medicine. *J. Hepatol.* 68 (3), 526–549. doi:10.1016/j.jhep.2017.09.016
- Helmkink, B. A., Reddy, S. M., Gao, J., Zhang, S., Basar, R., Thakur, R., et al. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* 577 (7791), 549–555. doi:10.1038/s41586-019-1922-8
- Hoshino, A., Costa-Silva, B., Shen, T. L., Rodrigues, G., Hashimoto, A., Tesic Mark, M., et al. (2015). Tumor exosome integrins determine organotropic metastasis. *Nature* 527 (7578), 329–335. doi:10.1038/nature15756
- Hydbring, P., Wang, Y., Fassl, A., Li, X., Matia, V., Otto, T., et al. (2017). Cell-cycle-targeting MicroRNAs as therapeutic tools against refractory cancers. *Cancer Cell* 31 (4), 576–590 e8. doi:10.1016/j.ccell.2017.03.004
- Iqbal, M. A., Arora, S., Prakasam, G., Calin, G. A., and Syed, M. A. (2019). MicroRNA in lung cancer carcinogenesis, pathways and therapeutic relevance. *Mol. Asp. Med.* 70, 3–20. doi:10.1016/j.mam.2018.07.003
- Kharaziha, P., Ceder, S., Li, Q., and Panaretakis, T. (2012). Tumor cell-derived exosomes: A message in a bottle. *Biochim. Biophys. Acta* 1826 (1), 103–111. doi:10.1016/j.bbcan.2012.03.006
- Li, C., Zhang, Z., Peng, E., and Peng, J. (2022). Role of an exosomes-related lncRNAs signature in tumor immune microenvironment of gastric cancer. *Front. Cell Dev. Biol.* 10, 873319. doi:10.3389/fcell.2022.873319
- Li, M. Y., Liu, L. Z., and Dong, M. (2021). Progress on pivotal role and application of exosome in lung cancer carcinogenesis, diagnosis, therapy and prognosis. *Mol. Cancer* 20 (1), 22. doi:10.1186/s12943-021-01312-y
- Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45 1, 255–268. doi:10.2307/2532051
- Lin, Y., Huang, K., Cai, Z., Chen, Y., Feng, L., Gao, Y., et al. (2022). A novel exosome-relevant molecular classification uncovers distinct immune escape mechanisms and genomic alterations in gastric cancer. *Front. Pharmacol.* 13, 884090. doi:10.3389/fphar.2022.884090
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2 (1), 56–67. doi:10.1038/s42256-019-0138-9
- Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501
- Mashouri, L., Yousefi, H., Aref, A. R., Ahadi, A. M., Molaie, F., and Alahari, S. K. (2019). Exosomes: Composition, biogenesis, and mechanisms in cancer metastasis and drug resistance. *Mol. Cancer* 18 (1), 75. doi:10.1186/s12943-019-0991-5
- Peinado, H., Zhang, H., Matei, I. R., Costa-Silva, B., Hoshino, A., Rodrigues, G., et al. (2017). Pre-metastatic niches: Organ-specific homes for metastases. *Nat. Rev. Cancer* 17 (5), 302–317. doi:10.1038/nrc.2017.6
- Petitprez, F., de Reynies, A., Keung, E. Z., Chen, T. W., Sun, C. M., Calderaro, J., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature* 577 (7791), 556–560. doi:10.1038/s41586-019-1906-8
- Pyzer, A. R., Cole, L., Rosenblatt, J., and Avigan, D. E. (2016). Myeloid-derived suppressor cells as effectors of immune suppression in cancer. *Int. J. Cancer* 139 (9), 1915–1926. doi:10.1002/ijc.30232
- Ren, R., Sun, H., Ma, C., Liu, J., and Wang, H. (2019). Colon cancer cells secrete exosomes to promote self-proliferation by shortening mitosis duration and activation of STAT3 in a hypoxic environment. *Cell Biosci.* 9, 62. doi:10.1186/s13578-019-0325-8
- Robbins, P. D., and Morelli, A. E. (2014). Regulation of immune responses by extracellular vesicles. *Nat. Rev. Immunol.* 14 (3), 195–208. doi:10.1038/nri3622
- Sautes-Fridman, C., Petitprez, F., Calderaro, J., and Fridman, W. H. (2019). Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* 19 (6), 307–325. doi:10.1038/s41586-019-0144-6
- Schulze, K., Nault, J. C., and Villanueva, A. (2016). Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *J. Hepatol.* 65 (5), 1031–1042. doi:10.1016/j.jhep.2016.05.035
- Sharonov, G. V., Serebrovskaya, E. O., Yuzhakova, D. V., Britanova, O. V., and Chudakov, D. M. (2020). B cells, plasma cells and antibody repertoires in the tumour microenvironment. *Nat. Rev. Immunol.* 20 (5), 294–307. doi:10.1038/s41577-019-0257-x
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., and Jemal, A. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- Tao, W. (2005). The mitotic checkpoint in cancer therapy. *Cell Cycle* 4 (11), 1495–1499. doi:10.4161/cc.4.11.2130
- Weaver, B. A., and Cleveland, D. W. (2005). Decoding the links between mitosis, cancer, and chemotherapy: The mitotic checkpoint, adaptation, and cell death. *Cancer Cell* 8 (1), 7–12. doi:10.1016/j.ccr.2005.06.011
- Whiteside, T. L. (2016). Tumor-derived exosomes and their role in tumor-induced immune suppression. *Vaccines (Basel)* 4 (4). doi:10.3390/vaccines4040035
- Wieckowski, E. U., Visus, C., Szajnik, M., Szczepanski, M. J., Storkus, W. J., and Whiteside, T. L. (2009). Tumor-derived microvesicles promote regulatory T cell expansion and induce apoptosis in tumor-reactive activated CD8+ T lymphocytes. *J. Immunol.* 183 (6), 3720–3730. doi:10.4049/jimmunol.0900970
- Wortzel, I., Dror, S., Kenific, C. M., and Lyden, D. (2019). Exosome-mediated metastasis: Communication from a distance. *Dev. Cell* 49 (3), 347–360. doi:10.1016/j.devcel.2019.04.011
- Wu, J., Zeng, D., Zhi, S., Ye, Z., Qiu, W., Huang, N., et al. (2021). Single-cell analysis of a tumor-derived exosome signature correlates with prognosis and immunotherapy response. *J. Transl. Med.* 19 (1), 381. doi:10.1186/s12967-021-03053-4
- Wu, Z. H., Li, C., Zhang, Y. J., and Lin, R. (2022). Bioinformatics study revealed significance of exosome transcriptome in hepatocellular carcinoma diagnosis. *Front. Cell Dev. Biol.* 10, 813701. doi:10.3389/fcell.2022.813701
- Yang, D., Zhang, W., Zhang, H., Zhang, F., Chen, L., Ma, L., et al. (2020). Progress, opportunity, and perspective on exosome isolation - efforts for efficient exosome-based theranostics. *Theranostics* 10 (8), 3684–3707. doi:10.7150/thno.41580
- Yuan, O., Lin, C., Wagner, J., Archard, J. A., Deng, P., Halmaj, J., et al. (2019). Exosomes derived from human primed mesenchymal stem cells induce mitosis and potentiate growth factor secretion. *Stem Cells Dev.* 28 (6), 398–409. doi:10.1089/scd.2018.0200
- Zhang, L., and Yu, D. (2019). Exosomes in cancer development, metastasis, and immunity. *Biochim. Biophys. Acta Rev. Cancer* 1871 (2), 455–468. doi:10.1016/j.bbcan.2019.04.004