Research on predicting the occurrence of hepatocellular carcinoma based on Notch signal-related genes using machine learning algorithms

Dingzhong Zhou
Affiliated Hospital (Clinical College)of Xiangnan University
Sujuan Cao
Affiliated Hospital (Clinical College) of Xiangnan University
Hui Xie (xieh612064@163.com)
Affiliated Hospital (Clinical College) of Xiangnan University

Research Article

Keywords: Hepatocellular carcinoma, Notch signal-related genes, Machine learning algorithms, classification and diagnosis model

Posted Date: March 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1418981/v1

Abstract

Background: Hepatocellular carcinoma (HCC) is a highly malignant malignant tumor, and is difficult to diagnose, treat and predict the prognosis. Notch signaling pathway can affect HCC. Therefore, our paper aimed to explore the prediction of the occurrence of hepatocellular carcinoma based on Notch signal-related genes using machine learning algorithms.

Methods: In our presented study, we downloaded HCC data from TCGA and GEO databases. Firstly, we used machine learning methods to screen the hub Notch signal-related genes. Then, machine learning classification was used to construct a prediction model for the classification and diagnosis of HCC cancer. In our presented study, we also used bioinformatics methods to explore the relationship between the expression of these hub genes in the HCC tumor immune microenvironment to further improve the reliablity of the model.

Results: After screening, we identified four hub genes of LAMA4, POLA2, RAD51 and TYMS, which were used as the final variables to construct the model. It was found that AdaBoostClassifie was the best algorithm for the classification and diagnosis model of HCC. The area under curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and F1 score of this model in the training set were 0.976, 0.881, 0.877, 0.977, 0.996, 0.500 and 0.932; respectively. The AUC, accuracy, sensitivity, specificity, PPV, NPV and F1 score in the testing set were 0.934, 0.863, 0.881, 0.886, 0.981, 0.489 and 0.926. The AUC in the external validation set was 0.934. In the immune microenvironment, immune cell infiltration played an important role in HCC and was related to the expression of 4 hub genes. Further researches found that patients in the low-risk group of HCC were more likely to have immune escape.

Conclusion: Notch signaling pathway were closely related to the occurrence and development of HCC. The HCC classification and diagnosis model established based on this had a high degree of reliability and stability.

Introduction

Primary liver cancer is one of the most common malignant tumors that seriously threaten human health in the world, and hepatocellular carcinoma (HCC) is the most common type of primary liver cancer. The latest report [1] released by the International Anti-Cancer Alliance shows that primary liver cancer is the fifth most likely malignant tumor for men and the seventh most likely malignant tumor for women. Currently, the global incidence of HCC is increasing year by year, and there are 1 million new cases of HCC each year [2]. The main treatment for HCC is surgical resection. Although surgical treatment may be effective for early HCC, the overall 5-year survival rate of patients is only 50–70% [3]. Moreover, up to 60–70% of patients experience tumor recurrence within 5 years after surgery, and the long-term prognosis after hepatectomy is poor [4]. Therefore, early diagnosis and early treatment of HCC are particularly

important for the prognosis of HCC, and it is necessary to develop a new model to predict the occurrence of HCC.

The Notch signaling pathway is a classic signaling pathway, and its family members are highly conserved in structure. With the in-depth study of the Notch signaling pathway, it is found that this pathway plays an important role in the occurrence and development of tumors [5]. It has been reported that the Notch signaling pathway can promote the development of cervical cancer cells, resulting in the formation of tumors [5]. Similarly, down regulation of Notch1 expression in pancreatic cancer can significantly inhibit the growth of pancreatic cancer cells, promote cell apoptosis, and stop the cell cycle at G0-G1 [6]. Gramantieri *et al.* found that the expression of Notch3 and Notch4 in liver cancer was significantly higher than that in adjacent tissues, Notch3 and Notch4 were also expressed in normal liver tissues and chronic hepatitis tissues, and Notch signaling pathway may participate in the invasion and metastasis of liver cancer [7]. However, it is unclear whether Notch signal-related genes (NSRGs) are related to the prognosis of HCC, and it is necessary to further explore the exact relationship between immune infiltrating cells in HCC microenvironment and NSRGs. With the rapid development of machine learning algorithms, we used it to study the above two problems and write this paper.

Materials And Methods

Data collection

In October 2021, we downloaded the data of 424 HCC cases (tumor tissue: 374 cases, normal tissues: 50 cases) from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/) as a training set and main research cohort. Then, we downloaded the HCC patient data (normal tissues:192, tumor tissues: 240) of the GSE36376 and platform GPL10558433 from the the Gene Expression Ommnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) as an external validation data set.

We downloaded 237 NSRGs from Molecular Signatures Database (MSigDB) (http://www.gsea-msigdb.org/gsea/msigdb) for research.

We used the R language to extract the expression matrix of NSRGs. All expression data were standardized by the Z-score processing (the mean value of the sample becomes 0, and the variance becomes 1). NSRGs were set as independent variables (feature), and normal samples/tumor samples as dependent variables (label) for the occurrence of HCC.

Differential expression analysis of NSRGs

We used the "limma" package of R language to select the differentially expressed NSRGs (DENSRGs) of HCC, with the criteria of |logFC|>1 and FDR<0.05 (FC: Fold Change, FDR: False Discovery Rate).

Identification of hub NSRGs

In the TCGA-HCC cohort, two machine learning algorithms of Least Absolute Shrinkage and Selection Operator (Lasso) and Support Vector Machine (SVM) were used to screen the important NSRGs of HCC.

Lasso is a kind of regression analysis algorithm, which selects variables while regularizing. Lasso was implemented by the "glmnet" package of R language with parameter settings of family:binomial, alpha:1, type, measure:deviance, nfolds:10.

SVM can express complex classification boundaries by combining with the kernel function. SVM was implemented by the "e1071, kernalb, caret" packages of R language with parameter settings of functions: careFuncs, method: cv, methods: svmRadial. In order to avoid the over-fitting of the model, we also performed univariate regression analysis of the gene in the selection of the feature genes using the "survival, survminer" packages of R language with the filter condition of P value < 0.05.

The intersecting genes of the three methods were identified as the final feature genes for further research as the variables of the classification and diagnosis model.

Establishment and verification of the prognostic model

The model of HCC classification and diagnosis was constructed based on the expression of core genes. In this study, five classification algorithms of machine learning classification algorithms, including XGBClassifier, LGBMClassifier, AdaBoostClassifier, MLPClassifier and SVM, were used to construct the the initial classification and diagnosis model.

30% of TCGA-HCC patients were randomly selected as the test set, the remaining 70% as the training set for the 10-fold cross-validation, and GEO-HCC patients as the external validation set were used to validate the model. Model evaluation indicators included area under curve (AUC), accuracy, sensitivity, positive predictive value (PPV), negative predictive value (NPV) and F1 score. TP, TN, FP and FN represented the number of true positive, true negative, false positive and false negative samples, respectively. Comprehensive evaluation of various indicators and selection of the best algorithm were used to build the model

$$Accuracy=(TP+TN)/(TP+TN+FP+FN)$$

$$Sensitivity=TP/(TP+FN)$$

$$F1=2TP/(2TP+FN+FP)$$

$$PPV= TP / (TP + FP)$$

$$NPV= TN / (TN + FN)$$

XGBClassifier was implemented based on the "xgboost 1.2.1" package of python language, and this model parameters were objective (optimized objective function): binary, logistic, learning_rate (learning rate): 0.3, max_depth (maximum tree depth): 4, min_child_weight (minimum bifurcation weight sum): 6,

and reg_lambda (L2 regularization coefficient): 1. The LGBMClassifier machine learning model was based on the "lightgbm 3.2.1" package of python language, and this model parameters were boosting_type (algorithm type): gbdt, learning_rate (learning rate): 0.001, max_depth (maximum tree depth): 1, n_estimators (maximum number of trees) : 5, and num_leaves (the maximum number of leaves): 5. AdaBoostClassifier was implemented based on the “sklearn 0.22.1” package of python language, and this model parameters were learning_rate (learning rate): 0.1, and n_estimators (number of single models): 50. MLPClassifier was implemented based on the “sklearn 0.22.1 ” package of python language, and this model parameters were activation (non-linear function): logistic, hidden_layer_sizes (hidden layer width): (30, 30), and max_iter (number of iterations): 10. SVM was implemented based on the “sklearn 0.22.1” package of python language, and this model parameters were C (regularization factor): 0.1, kernel (core type): rbf, tol (convergence measure): 0.001.

Hub gene expression analysis

In order to further analyze the relationship between participating variables (hub genes) and HCC. The Mann–Whitney test was used to compare the expression levels of hub genes between tumor group and normal group. Pearson correlation was used to calculated the correlation of risk genes.

Immune cell infiltration analysis

The CIBERSORT algorithm was used to evaluate the infiltration of 22 immune cells in the TCGA-HCC cohort. Then, we compared the distribution of the 22 immune cells between normal group and tumor group. Spearman correlation analysis was performed between 22 immune cells and hub genes. At the end, the risk of patients was scored according to the gene expression of the selected variables in the model. The patients were divided into the high-risk group and the low-risk group by the median value of the risk score, and then analyzed for immunotherapy responsiveness. The risk score was calculated as the sum of the predicted values weighted by the Lasso coefficient, including all risk genes. The Tumor Immune Dysfunction and Exclusion (TIDE) tool (http://tide.dfci.harvard.edu/) was used to predict immunotherapy responsiveness.

Results

The flowchart of this study is shown in Fig. 1.

Differential expression genes analysis

67 DENSRGs were identified with with |logFC|>1 and FDR < 0.05, of which 4 genes were down-regulated, and 63 genes were up-regulated (Supplementary table 1).

Identification of hub genes

The Lasso algorithm analyzed the DENSRGs of the TCGA-HCC cohort to select key feature genes and determine the optimal value λ with the smallest mean square error through 10-fold cross-validation (Fig.

2A). When λ was 0.006, 17 feature genes were screened out (Fig. 2B). The main idea of the SVM algorithm is for the two classification problem, find a hyperplane and divide the two categories to ensure the minimum classification error rate. SVM analysis showed that 37 genes were closely related to HCC (Fig. 2C). HCC univariate regression analysis found that there were 18 NSRGs related to the survival of HCC (Fig. 2D). The intersection of the three methods finally resulted in four hub genes: LAMA4, POLA2, RAD51 and TYMS (Fig. 2E).

In our presented study, the classification and diagnosis model of HCC was constructed based on the expression levels of LAMA4, POLA2, RAD51 and TYMS (Fig. 3A). We chosen the best among five machine learning classification algorithms of XGBClassifier, LGBMClassifier, AdaBoostClassifier, MLPClassifier, and SVM by the metrics of AUC, Accuracy, Sensitivity, PPV, NPV, F1 score. The best performer in the training set was AdaBoostClassifier (Fig. 3B), and the corresponding scores in the training set in each evaluation standard were: AUC: 0.976, Accuracy: 0.881, Sensitivity: 0.877, Specificity: 0.977, PPV: 0.996, NPV : 0.500, and F1 score: 0.932 (Table 1). The best performer in the testing set was also AdaBoostClassifier (Fig. 3C), and the corresponding scores in the testing set in each evaluation standard were: AUC: 0.934, Accuracy: 0.863, Sensitivity: 0.881, Specificity: 0.886, PPV: 0.981, NPV : 0.489, and F1 score: 0.926 (Table 2). The results in the training set were consistent with those in the testing set, and AdaBoostClassifier was considered as the best model.

Table 1
Performance of the classification and diagnosis models based on 5 machine learning algorithms in the training set

Model	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
XGBClassifier	0.973 ± 0.007	0.893 ± 0.030	0.890 ± 0.035	0.959 ± 0.048	0.994 ± 0.006	0.529 ± 0.094	0.939 ± 0.018
LGBMClassifier	0.731 ± 0.027	0.120 ± 0.007	0.957 ± 0.021	0.505 ± 0.073	Na	0.120 ± 0.007	Na
AdaBoostClassifier	0.976 ± 0.007	0.881 ± 0.020	0.877 ± 0.024	0.977 ± 0.031	0.996 ± 0.005	0.500 ± 0.057	0.932 ± 0.013
MLPClassifier	0.822 ± 0.061	0.705 ± 0.090	0.691 ± 0.107	0.832 ± 0.058	0.968 ± 0.010	0.285 ± 0.077	0.802 ± 0.073
SVM	0.908 ± 0.016	0.873 ± 0.011	0.882 ± 0.012	0.840 ± 0.037	0.976 ± 0.006	0.480 ± 0.036	0.926 ± 0.007

Note: hepatocellular carcinoma: HCC; area under curve: AUC; positive predictive value: PPV; negative predictive value: NPV; Support Vector Machine: SVM

Table 2
Performance of the classification and diagnosis models based on 5 machine learning algorithms in the testing set

Model	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
XGBClassifier	0.901 ± 0.048	0.844 ± 0.065	0.853 ± 0.081	0.854 ± 0.105	0.975 ± 0.019	0.470 ± 0.114	0.907 ± 0.046
LGBMClassifier	0.699 ± 0.051	0.113 ± 0.017	0.946 ± 0.036	0.451 ± 0.124	Na	0.113 ± 0.017	Na
AdaBoostClassifier	0.934 ± 0.033	0.863 ± 0.059	0.881 ± 0.077	0.886 ± 0.090	0.981 ± 0.016	0.489 ± 0.128	0.926 ± 0.042
MLPClassifier	0.867 ± 0.085	0.788 ± 0.148	0.785 ± 0.179	0.865 ± 0.118	0.980 ± 0.015	0.429 ± 0.195	0.859 ± 0.120
SVM	0.866 ± 0.064	0.867 ± 0.062	0.886 ± 0.068	0.791 ± 0.118	0.970 ± 0.017	0.487 ± 0.127	0.925 ± 0.040

Note: hepatocellular carcinoma: HCC; area under curve: AUC; positive predictive value: PPV; negative predictive value: NPV; Support Vector Machine: SVM

Validation model

We used GEO-HCC cohort data to verify the classification and diagnosis model constructed by the AdaBoostClassifier method. All patients were verified as AUC = 0.940 in the HCC tumor and normal tissue classification model (Fig. 3D).

Gene expression analysis

The four risk genes of LAMA4, POLA2, RAD51 and TYMS showed significant differences in the expression of normal and tumor tissues, and the four genes showed high expression in tumor tissues (Fig. 4A). We found that POLA2 and TYMS were highly positively correlated with RAD51, while LAM4 was highly negatively correlated with RAD51(Fig. 4B).

In the analysis of the difference in immune cell infiltration between normal and tumor tissues, Tregs ($p < 0.001$), Monocytes ($p = 0.024$), Macrophages ($p < 0.001$) and Neutrophils ($p = 0.002$) showed differences (Fig. 4C and 4D). The correlation analysis between immune infiltrating cells suggested that Macrophages and Tregs were negatively correlated with T cells (Fig. 5A). The correlation analysis between risk genes and immune cell infiltration subtypes found that NK cells activated had no correlation with LAMA4, POLA2, RAD51 and TYMS. However, almost all of the other immune cell infiltration subtypes showed varying degrees of correlation with risk genes (Fig. 5B-5E). We used the public website http://tide.dfci.harvard.edu to perform analysis of Tumor Immune Dysfunction and Evolution (TIDE),

Exclusion, Dysfunction and MSI immunotherap of HCC, and found that HCC patients showed differences in TIDE, Exclusion, and Dysfunction between the high-risk group and the low-risk group (Fig. 6A-6D).

Discussion

The mortality of HCC ranks second among all kinds of cancers, and the new cases of HCC in China each year accounts for more than half of new cases around the world [8]. The treatment of HCC is affected by liver function, nodule size, metastasis and age. Notch signaling pathway as a classic signaling pathway can regulate the occurrence of tumor cells [9]. There is evidence that the Notch signaling pathway is extraordinarily active in HCC [10]. At present, surgical resection is still the main option for HCC, but its recurrence risk is very high. With the help of machine learning methods, the prediction accuracy of early HCC can be improved, thereby further improving the treatment outcome of HCC patients.

In our presented study, we firstly used Lasso and SVM algorithms combined with univariate survival analysis to study the differential expression matrix of NSRGs in the TCGA-HCC cohort. Then, after comparing the results of 5 machine learning classification algorithms, we finally decided to use AdaBoostClassifier to establish the HCC classification and diagnosis model. The HCC classification and diagnosis model established in our presented study had AUC of 0.976 ± 0.007 in the training set and 0.934 ± 0.033 in the test set. In the external data test of the GEO-HCC cohort, this model had AUC of 0.934, Accuracy of 0.863, Sensitivity of 0.881, Specificity of 0.886, PRV of 0.981, NPV of 0.489, and F1 score of 0.926. These results showed that the classification and diagnosis model established in our presented study might be highly reliable. Duan *et al.* [11] used the BP neural network to establish an auxiliary diagnosis model of lung cancer with genes of P16, RASSF1A and FHIT, and the AUC of this model was 0.76. Sherafatian and Arjand *et al.* [12] used decision tree (DT) method and lung adenocarcinoma and lung squamous cell carcinoma datasets from TCGA to construct a diagnostic model for classifying the sample as lung squamous cell carcinoma, and the AUC of this model was 0.916. Yu *et al.* [13] used CT imaging data combined with machine learning methods to diagnose lung cancer patients and determine their pathological stages, and the AUC of the final model ranged from 0.69 to 1.00. It can be seen that the machine learning method can solve the classification problem of lung cancer very well. In our presented study, the HCC diagnosis model established by AdaBoostClassifier had AUC of 0.976 ± 0.007 in the training set, 0.934 ± 0.033 in the testing set, and 0.932 in the external validation set, which suggested that our presented study has certain advantages and reliability compared with similar models.

Studies have shown that the downregulation of LAMA4 expression can inhibit proliferation and migration of breast cancer, renal cell carcinoma, gastric cancer and ovarian cancer[14–16]. Our study showed that LAMA4 was highly expressed in HCC tumor tissues. Considering that LAMA4 is closely related to the migration of cancer cells and tumor progression in a series of tumors, the the latest research describes LAMA4 as "oncolaminin" [17]. The crosstalk between Notch and TGF-β1 has been reported many times. It has been reported that LAMA4 could affect the level of Notch ligand and its receptor by regulating TGF-β1 [18], thereby inducing the expression of some key proteins related to the occurrence and development of HCC. Cir_POLA2 has been reported as an oncogene of lung cancer [19]. It has been found that

overexpression of Cir_POLA2 can promote the proliferation of acute myeloid leukemia (AML) cells [20]. Circ_POLA2 may upregulate the G protein subunit beta 1 (Notch pathway related molecules) by serving as an endogenous competing RNA for miR-326.14 [21]. Guanine nucleotide regulatory protein (G protein) is the core of normal liver cell function and is related to the occurrence and progression of liver disease. It was reported that the G protein family was involved in the development of HCC [22]. In our presented study, when we performed the comparison between liver tumor tissue and paracancerous tissue, it was found POLA2 was relatively lower in the normal paracancerous tissues. As we know, the DNA repair protein RAD51 mainly plays an important role in maintaining the genome stability and regulating cell life cycle. It has been reported that the DNA repair system in most HCC cells is extraordinarily active, resulting in poor therapeutic effect of HCC [23]. RAD51 is a key protein for DNA double-strand repair. Highly expressed RAD51 promotes the repair of HCC [24, 25]. Chen Q *et al.* [26] have reported that in female ovarian cancer, knocking down the expression of RAD51 can significantly reduce the proliferation rate of ovarian cancer cells. It has bee found that the high expression of RAD51 is related to the higher pathological grade and clinical stage of HCC, and it is an independent risk factor affecting the overall survival and prognosis of HCC [27]. TYMS is the key rate-limiting enzyme that controls the synthesis of dTMP. The synthesis of dTMP is closely related to functions such as DNA synthesis, replication and repair. Therefore, TYMS is currently considered to be the next anti-tumor target which is most likely to be successfully developed [28]. Studies have shown that the activity of thymidylate synthase in many patients with malignant tumors is significantly higher than that in normal tissues [28]. TYMS can regulate the growth of tumor cells by affecting the expression and expression cycle of P53, so TYMS is related to the proliferation status of malignant tumors [29]. Studies have shown that in most of the tumor cells with growth advantages TYMS are over-expressed, and the higher the expression of TYMS is, the worse the prognosis of patients is [29]. A study showed that the positive expression rate of TYMS in liver cancer tissues was significantly higher than that in the control group and adjacent tissues, and the high expression of TYMS indicated that the tumor was more aggressive [30].

Immune infiltrating cells are an important part of the tumor microenvironment (TME) which are closely related to the progression of HCC [31]. In this study, we used the CIBERSORT algorithm to evaluate patients' immune cell infiltration. Further analysis found that there were differences in T cells regulatory (Tregs), Monocytes, Macrophages M0 and Neutrophils between normal tissues and tumor tissues ($P < 0.05$). The proportion of Tregs and Macrophages M0 in tumor tissues is higher than that in normal tissues, and Tregs are strongly positively correlated with Macrophages M0, while the proportion of Monocytes and Neutrophils in tumor tissues is lower than that in normal tissues, and Monocytes are strongly negatively correlated with Neutrophils. The increase of Macrophages M0 is significantly correlated with OS and tumor stage of HCC [32]. Macrophages M0 can stimulate the production of TAM and Kupffer cells in the presence of carcinogenic factors, thereby inhibiting the progression of HCC caused by immunity [33]. This may be related to the malignant behavior of the highly expressed genes from our research and analysis. The correlation analysis of our presented study also confirmed the significant positive correlation between LAMA4, RAD51 and TYMS and Macrophages M0. It has been reported that TGF-β1 is strongly positively correlated with macrophage in HCC [34]. Macrophages M0 can

secrete a large amount of TGF-β1 [34]. In our study, it was found that LAMA4 affected the progress of HCC by regulating TGF-β1. This also shows the correctness of this research. Under normal circumstances, Tregs inhibit the anti-autoimmune response and play an important role in balancing immune tolerance and inflammation. It has been reported that the upregulation of Tregs is a predictor of adverse outcomes of HCC patients [35]. It has been reported Tregs promote the migration and invasion of liver cancer cells through epithelial-to-mesenchymal transition (EMT) induced by TGF-β1 [36]. Neutrophils are the most common white blood cells in the circulation. They play an important role in host defense, immune regulation, and tissue damage. Neutrophils are considered to be one of the first immune cells that enter the tumor microenvironment and interact with cancer cells. Thus it plays an important role in the progression of cancer. Our research findings are similar to other studies. The content of neutrophils in normal tissues is high, and it is significantly negatively correlated with LAMP4, RAD51 and TYMS. In our presented study, we found RAD51 was strongly positively correlated with TYMS, and this interrelationship further strengthens the immune function of Neutrophils.

The liver has a special population of immunosuppressive cells, which can avoid liver damage caused by autoimmunity and chronic inflammation under normal physiological conditions. But for patients with liver cancer, these special cells can cause tumor immune escape and promote disease progression. Our presented study found that TIDE, Exclusion and Dysfunction in the low-risk population of HCC were all higher than those in the high-risk population of HCC ($P < 0.05$). This showed that immune escape was prone to occur in the low-risk population. The previous analysis show Tregs cells are abundant in HCC tumors, and are a subset of CD4 + T cells, a type of lymphocytes with high immunosuppressive properties [37]. They suppress the immune response by inhibiting CD8 + T cell effector functions, and directly promote tumor escape through a variety of contact-dependent and non-contact mechanisms [38]. In HCC, neutrophils can recruit macrophages and Tregs into HCC by releasing cytokines, thereby promoting tumor progression and developing resistance to sorafenib[39].

There are some limitation in our study. The biological functions of LAMA4, POLA2, RAD51 and TYMS need to be further explored by experiments. The construction and validation of our established model are only based on the public databases, and thus it is necessary to use more clinical research data to further validate clinical efficacy of this model.

In conclusion, the classification and diagnosis model of HCC based on NSRGs in our presented study showed that Notch signal was related to the occurrence and development of HCC. The classification diagnosis model might effectively distinguish HCC from the healthy patients. This study also suggested the role of immune infiltrating cells in the occurrence and development of HCC. Our findings provide a new method for the diagnosis and prognosis of HCC from the perspective of immune cell infiltration.

Declarations

Acknowledgements

Acknowledgments to the TCGA and ICGC databases for providing researchable patient data.

Funding: This study was supported by:

1. Science and Technology Funding Project of Hunan Province China (No.2019JJ80073)

2. Key Laboratory of Tumor Precision Medicine, Hunan colleges and Universities Project (2019-379)

Availability of data and materials

Not applicable.

Authors' contributions

Dingzhong Zhou and Sujuan Cao designed the study, searched, analyzed and interpreted the literature and was a major contributor in writing the manuscript. Xie Hui designed the study and revised the manuscript.

Conflict of interest statement and Consent for publication

The authors have no ethical, legal and financial conflicts related to the article. All authors read and approved the manuscript to publish.

References

1. Tseng HC, Xiong W, Badeti S, Yang Y, Ma M, Liu T, Ramos CA, Dotti G, Fritzky L, Jiang JG, Yi Q, Guarrera J, Zong WX, Liu C, Liu D. Efficacy of anti-CD147 chimeric antigen receptors targeting hepatocellular carcinoma. Nat Commun. 2020 Sep 23;11(1):4810. doi: 10.1038/s41467-020-18444-2. PMID: 32968061; PMCID: PMC7511348.
2. Myers S, Neyroud-Caspar I, Spahr L, Gkouvatsos K, Fournier E, Giostra E, Magini G, Frossard JL, Bascaron ME, Vernaz N, Zampaglione L, Negro F, Goossens N. NAFLD and MAFLD as emerging causes of HCC: A populational study. JHEP Rep. 2021 Jan 19;3(2):100231. doi: 10.1016/j.jhepr.2021.100231. PMID: 33748726; PMCID: PMC7957147.
3. Wei L, Lee D, Law CT, Zhang MS, Shen J, Chin DW, Zhang A, Tsang FH, Wong CL, Ng IO, Wong CC, Wong CM. Genome-wide CRISPR/Cas9 library screening identified PHGDH as a critical driver for Sorafenib resistance in HCC. Nat Commun. 2019 Oct 15;10(1):4681. doi: 10.1038/s41467-019-12606-7. PMID: 31615983; PMCID: PMC6794322.
4. Lin KT, Ma WK, Scharner J, Liu YR, Krainer AR. A human-specific switch of alternatively spliced AFMID isoforms contributes to TP53 mutations and tumor recurrence in hepatocellular carcinoma. Genome Res. 2018 Feb 15;28(3):275–84. doi: 10.1101/gr.227181.117. Epub ahead of print. PMID: 29449409; PMCID: PMC5848607.
5. Yue Y, Zhou K, Li J, Jiang S, Li C, Men H. MSX1 induces G0/G1 arrest and apoptosis by suppressing Notch signaling and is frequently methylated in cervical cancer. Onco Targets Ther. 2018 Aug 10;11:4769–4780. doi: 10.2147/OTT.S165144. PMID: 30127625; PMCID: PMC6091477.

6. Khan F, Pandey P, Jha NK, Khalid M, Ojha S. Rutin Mediated Apoptotic Cell Death in Caski Cervical Cancer Cells via Notch-1 and Hes-1 Downregulation. Life (Basel). 2021 Jul 28;11(8):761. doi: 10.3390/life11080761. PMID: 34440505; PMCID: PMC8400226.
7. Gramantieri L, Giovannini C, Lanzi A, Chieco P, Ravaioli M, Venturi A, Grazi GL, Bolondi L. Aberrant Notch3 and Notch4 expression in human hepatocellular carcinoma. Liver Int. 2007 Sep;27(7):997–1007. doi: 10.1111/j.1478-3231.2007.01544.x. PMID: 17696940.
8. Chen J, Wang L, Wang F, Liu J, Bai Z. Genomic Identification of RNA Editing Through Integrating Omics Datasets and the Clinical Relevance in Hepatocellular Carcinoma. Front Oncol. 2020 Feb 14;10:37. doi: 10.3389/fonc.2020.00037. PMID: 32117713; PMCID: PMC7033493.
9. Kumar V, Vashishta M, Kong L, Wu X, Lu JJ, Guha C, Dwarakanath BS. The Role of Notch, Hedgehog, and Wnt Signaling Pathways in the Resistance of Tumors to Anticancer Therapies. Front Cell Dev Biol. 2021 Apr 22;9:650772. doi: 10.3389/fcell.2021.650772. PMID: 33968932; PMCID: PMC8100510.
10. Kim BR, Na YJ, Kim JL, Jeong YA, Park SH, Jo MJ, Jeong S, Kang S, Oh SC, Lee DH. RUNX3 suppresses metastasis and stemness by inhibiting Hedgehog signaling in colorectal cancer. Cell Death Differ. 2020 Feb;27(2):676–694. doi: 10.1038/s41418-019-0379-5. Epub 2019 Jul 5. PMID: 31278361; PMCID: PMC7205901.
11. Duan X, Yang Y, Tan S, Wang S, Feng X, Cui L, Feng F, Yu S, Wang W, Wu Y. Application of artificial neural network model combined with four biomarkers in auxiliary diagnosis of lung cancer. Med Biol Eng Comput. 2017 Aug;55(8):1239–1248. doi: 10.1007/s11517-016-1585-7. Epub 2016 Oct 20. PMID: 27766520.
12. Sherafatian M, Arjmand F. Decision tree-based classifiers for lung cancer diagnosis and subtyping using TCGA miRNA expression data. Oncol Lett. 2019 Aug;18(2):2125–2131. doi: 10.3892/ol.2019.10462. Epub 2019 Jun 10. PMID: 31423286; PMCID: PMC6607095.
13. Yu L, Tao G, Zhu L, Wang G, Li Z, Ye J, Chen Q. Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. BMC Cancer. 2019 May 17;19(1):464. doi: 10.1186/s12885-019-5646-9. PMID: 31101024; PMCID: PMC6525347.
14. Ross JB, Huh D, Noble LB, Tavazoie SF. Identification of molecular determinants of primary and metastatic tumour re-initiation in breast cancer. Nat Cell Biol. 2015 May;17(5):651–64. doi: 10.1038/ncb3148. Epub 2015 Apr 13. PMID: 25866923; PMCID: PMC4609531.
15. Zheng B, Qu J, Ohuchida K, Feng H, Chong SJF, Yan Z, Piao Y, Liu P, Sheng N, Eguchi D, Ohtsuka T, Mizumoto K, Liu Z, Pervaiz S, Gong P, Nakamura M. Erratum: LAMA4 upregulation is associated with high liver metastasis potential and poor survival outcome of Pancreatic Cancer: Erratum. Theranostics. 2021 Nov 10;11(20):10171–10172. doi: 10.7150/thno.68023. Erratum for: Theranostics. 2020 Aug 13;10(22):10274–10289. PMID: 34815810; PMCID: PMC8581420.
16. Wang T, Jiang R, Yao Y, Qian L, Zhao Y, Huang X. Identification of endometriosis-associated genes and pathways based on bioinformatic analysis. Medicine (Baltimore). 2021 Jul 9;100(27):e26530. doi: 10.1097/MD.0000000000026530. PMID: 34232189; PMCID: PMC8270630.

17. Mukai C, Choi E, Sams KL, Klampen EZ, Anguish L, Marks BA, Rice EJ, Wang Z, Choate LA, Chou SP, Kato Y, Miller AD, Danko CG, Coonrod SA. Chromatin run-on sequencing analysis finds that ECM remodeling plays an important role in canine hemangiosarcoma pathogenesis. BMC Vet Res. 2020 Jun 22;16(1):206. doi: 10.1186/s12917-020-02395-3. PMID: 32571313; PMCID: PMC7310061.
18. Monteiro R, Pinheiro P, Joseph N, Peterkin T, Koth J, Repapi E, Bonkhofer F, Kirmizitas A, Patient R. Transforming Growth Factor β Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells. Dev Cell. 2016 Aug 22;38(4):358 – 70. doi: 10.1016/j.devcel.2016.06.024. Epub 2016 Aug 4. PMID: 27499523; PMCID: PMC4998007.
19. Fang X, Wang J, Chen L, Zhang X. circRNA circ_POLA2 increases microRNA-31 methylation to promote endometrial cancer cell proliferation. Oncol Lett. 2021 Nov;22(5):762. doi: 10.3892/ol.2021.13023. Epub 2021 Sep 6. PMID: 34539866; PMCID: PMC8436404.
20. Li H, Bi K, Feng S, Wang Y, Zhu C. CircRNA circ_POLA2 is Upregulated in Acute Myeloid Leukemia (AML) and Promotes Cell Proliferation by Suppressing the Production of Mature miR-34a. Cancer Manag Res. 2021 May 5;13:3629–3637. doi: 10.2147/CMAR.S281690. PMID: 33981162; PMCID: PMC8107013.
21. Cao Y, Li J, Jia Y, Zhang R, Shi H. CircRNA circ_POLA2 Promotes Cervical Squamous Cell Carcinoma Progression via Regulating miR-326/GNB1. Front Oncol. 2020 Jul 16;10:959. doi: 10.3389/fonc.2020.00959. PMID: 32766125; PMCID: PMC7381119.
22. Kimura T, Pydi SP, Pham J, Tanaka N. Metabolic Functions of G Protein-Coupled Receptors in Hepatocytes-Potential Applications for Diabetes and NAFLD. Biomolecules. 2020 Oct 15;10(10):1445. doi: 10.3390/biom10101445. PMID: 33076386; PMCID: PMC7602561.
23. Evert M, Frau M, Tomasi ML, Latte G, Simile MM, Seddaiu MA, Zimmermann A, Ladu S, Staniscia T, Brozzetti S, Solinas G, Dombrowski F, Feo F, Pascale RM, Calvisi DF. Deregulation of DNA-dependent protein kinase catalytic subunit contributes to human hepatocarcinogenesis development and has a putative prognostic value. Br J Cancer. 2013 Nov 12;109(10):2654-64. doi: 10.1038/bjc.2013.606. Epub 2013 Oct 17. PMID: 24136149; PMCID: PMC3833205.
24. Chen CC, Chen CY, Wang SH, Yeh CT, Su SC, Ueng SH, Chuang WY, Hsueh C, Wang TH. Melatonin Sensitizes Hepatocellular Carcinoma Cells to Chemotherapy Through Long Non-Coding RNA RAD51-AS1-Mediated Suppression of DNA Repair. Cancers (Basel). 2018 Sep 10;10(9):320. doi: 10.3390/cancers10090320. PMID: 30201872; PMCID: PMC6162454.
25. Chen CC, Chen CY, Wang SH, Yeh CT, Su SC, Ueng SH, Chuang WY, Hsueh C, Wang TH. Melatonin Sensitizes Hepatocellular Carcinoma Cells to Chemotherapy Through Long Non-Coding RNA RAD51-AS1-Mediated Suppression of DNA Repair. Cancers (Basel). 2018 Sep 10;10(9):320. doi: 10.3390/cancers10090320. PMID: 30201872; PMCID: PMC6162454.
26. Chen Q, Cai D, Li M, Wu X. The homologous recombination protein RAD51 is a promising therapeutic target for cervical carcinoma. Oncol Rep. 2017 Aug;38(2):767–774. doi: 10.3892/or.2017.5724. Epub 2017 Jun 15. PMID: 28627709; PMCID: PMC5561999.

27. Pa CZ, Zhang GQ, Wu SM. Relationship between RAD51 expression and proliferation, invasion and prognosis of hepatocellular carcinoma. J Med Res. 2021;50(09):127–132. DOI: 10.11969/j.issn.1673-548X.2021.09.029. (in Chinese)
28. Reich S, Nguyen CDL, Has C, Steltgens S, Soni H, Coman C, Freyberg M, Bichler A, Seifert N, Conrad D, Knobbe-Thomsen CB, Tews B, Toedt G, Ahrends R, Medenbach J. A multi-omics analysis reveals the unfolded protein response regulon and stress-induced resistance to folate-based antimetabolites. Nat Commun. 2020 Jun 10;11(1):2936. doi: 10.1038/s41467-020-16747-y. PMID: 32522993; PMCID: PMC7287054.
29. Botcheva K, McCorkle SR, McCombie WR, Dunn JJ, Anderson CW. Distinct p53 genomic binding patterns in normal and cancer-derived human cells. Cell Cycle. 2011 Dec 15;10(24):4237-49. doi: 10.4161/cc.10.24.18383. Epub 2011 Dec 15. PMID: 22127205; PMCID: PMC3272258.
30. Gandhi M, Groß M, Holler JM, Coggins SA, Patil N, Leupold JH, Munschauer M, Schenone M, Hartigan CR, Allgayer H, Kim B, Diederichs S. The lncRNA lincNMR regulates nucleotide metabolism via a YBX1 - RRM2 axis in cancer. Nat Commun. 2020 Jun 25;11(1):3214. doi: 10.1038/s41467-020-17007-9. PMID: 32587247; PMCID: PMC7316977.
31. Wu TJ, Chang SS, Li CW, Hsu YH, Chen TC, Lee WC, Yeh CT, Hung MC. Severe Hepatitis Promotes Hepatocellular Carcinoma Recurrence via NF-κB Pathway-Mediated Epithelial-Mesenchymal Transition after Resection. Clin Cancer Res. 2016 Apr 1;22(7):1800-12. doi: 10.1158/1078-0432.CCR-15-0780. Epub 2015 Dec 11. PMID: 26655845; PMCID: PMC4818680.
32. Huang R, Liu J, Li H, Zheng L, Jin H, Zhang Y, Ma W, Su J, Wang M, Yang K. Identification of Hub Genes and Their Correlation With Immune Infiltration Cells in Hepatocellular Carcinoma Based on GEO and TCGA Databases. Front Genet. 2021 Apr 30;12:647353. doi: 10.3389/fgene.2021.647353. PMID: 33995482; PMCID: PMC8120231.
33. Chen W, Zhang X, Bi K, Zhou H, Xu J, Dai Y, Diao H. Comprehensive Study of Tumor Immune Microenvironment and Relevant Genes in Hepatocellular Carcinoma Identifies Potential Prognostic Significance. Front Oncol. 2020 Sep 24;10:554165. doi: 10.3389/fonc.2020.554165. PMID: 33072579; PMCID: PMC7541903.
34. Huang Y, Lu J, Xu Y, Xiong C, Tong D, Hu N, Yang H. Xiaochaihu decorction relieves liver fibrosis caused by Schistosoma japonicum infection via the HSP47/TGF-β pathway. Parasit Vectors. 2020 May 14;13(1):254. doi: 10.1186/s13071-020-04121-2. PMID: 32410640; PMCID: PMC7227055.
35. Zhang AB, Qian YG, Zheng SS. Prognostic significance of regulatory T lymphocytes in patients with hepatocellular carcinoma. J Zhejiang Univ Sci B. 2016 Dec.;17(12):984–991. doi: 10.1631/jzus.B1600264. PMID: 27921403; PMCID: PMC5172602.
36. Pezone A, Taddei ML, Tramontano A, Dolcini J, Boffo FL, De Rosa M, Parri M, Stinziani S, Comito G, Porcellini A, Raugei G, Gackowski D, Zarakowska E, Olinski R, Gabrielli A, Chiarugi P, Avvedimento EV. Targeted DNA oxidation by LSD1-SMAD2/3 primes TGF-β1/ EMT genes for activation or repression. Nucleic Acids Res. 2020 Sep 18;48(16):8943–8958. doi: 10.1093/nar/gkaa599. PMID: 32697292; PMCID: PMC7498341.

37. Chandrasekar AP, Cummins NW, Badley AD. The Role of the BCL-2 Family of Proteins in HIV-1 Pathogenesis and Persistence. Clin Microbiol Rev. 2019 Oct 30;33(1):e00107-19. doi: 10.1128/CMR.00107-19. PMID: 31666279; PMCID: PMC6822993.
38. Wang X, Waschke BC, Woolaver RA, Chen SMY, Chen Z, Wang JH. HDAC inhibitors overcome immunotherapy resistance in B-cell lymphoma. Protein Cell. 2020 Jul;11(7):472–482. doi: 10.1007/s13238-020-00694-x. Epub 2020 Mar 11. PMID: 32162275; PMCID: PMC7305292.
39. Coy S, Rashid R, Lin JR, Du Z, Donson AM, Hankinson TC, Foreman NK, Manley PE, Kieran MW, Reardon DA, Sorger PK, Santagata S. Multiplexed immunofluorescence reveals potential PD-1/PD-L1 pathway vulnerabilities in craniopharyngioma. Neuro Oncol. 2018 Jul 5;20(8):1101–1112. doi: 10.1093/neuonc/noy035. PMID: 29509940; PMCID: PMC6280314.

Figures

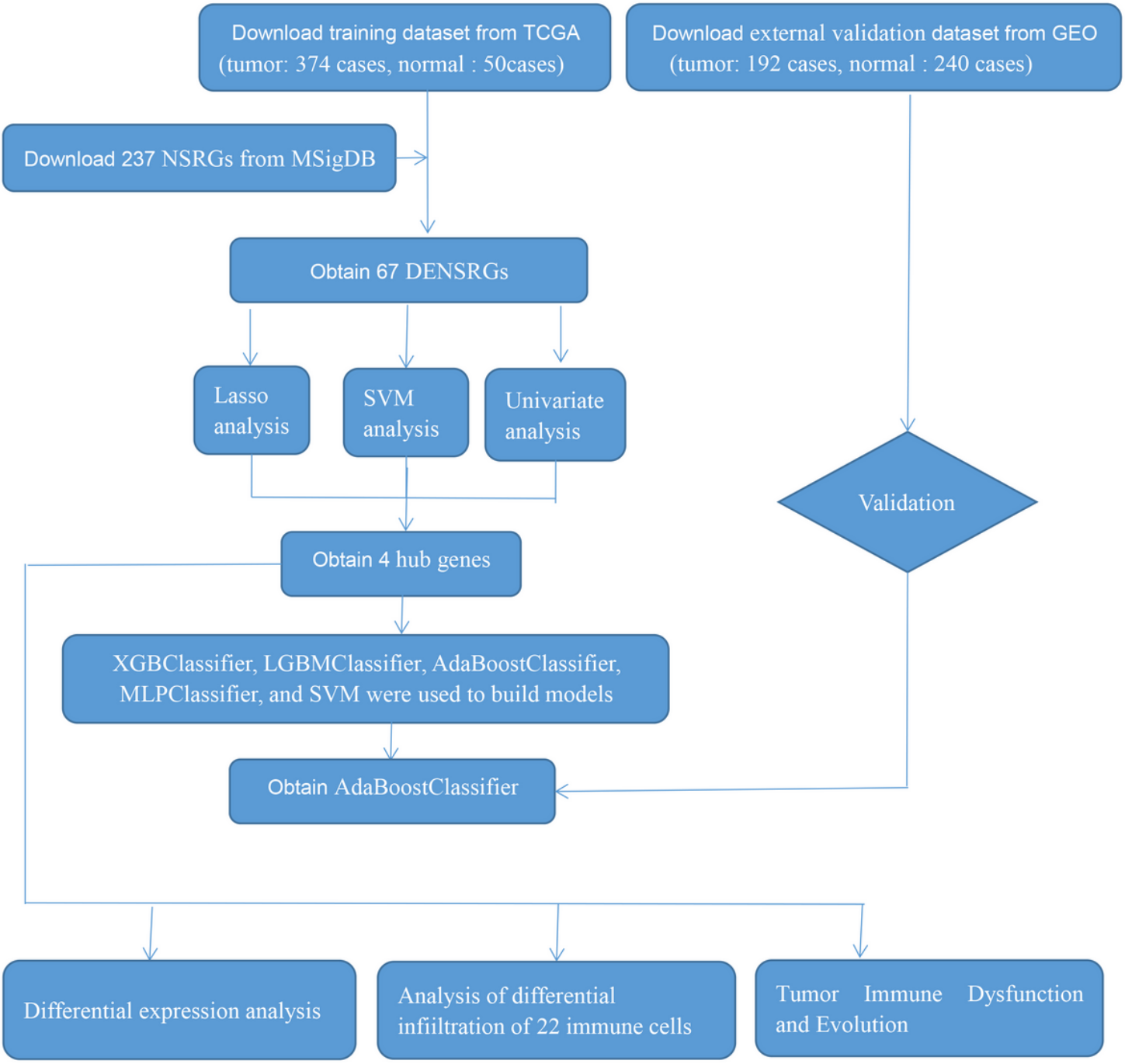

Figure 1

The flow-chart of the whole study.

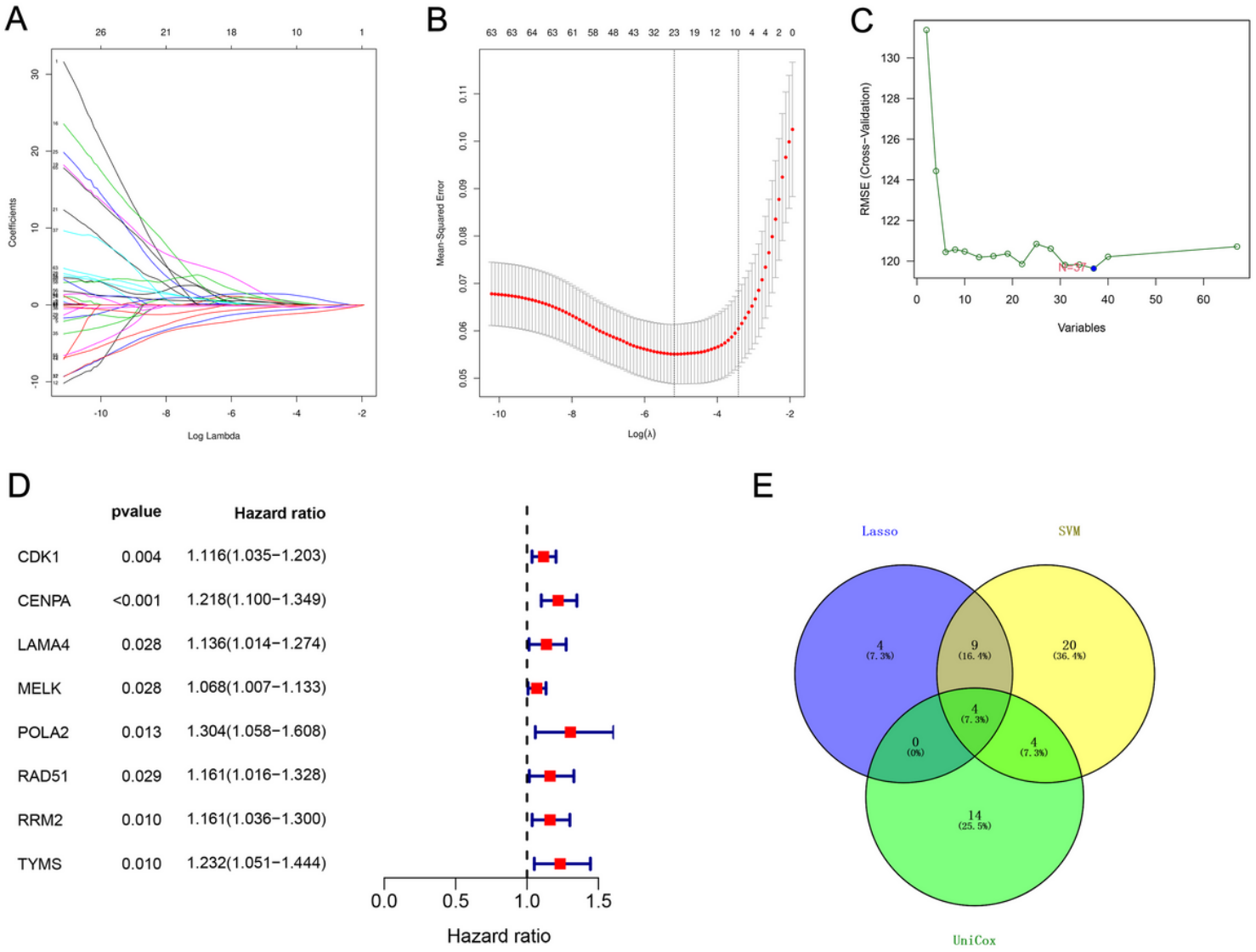

Figure 2

Variables screening process. (A) Trend graph of LASSO coefficients; (B) Partial likelihood deviation map; (C) Using support vector machine (SVM)-RFE feature selection; (D) Univariate Cox regression analysis; (E) Venn diagram of overlapping genes.

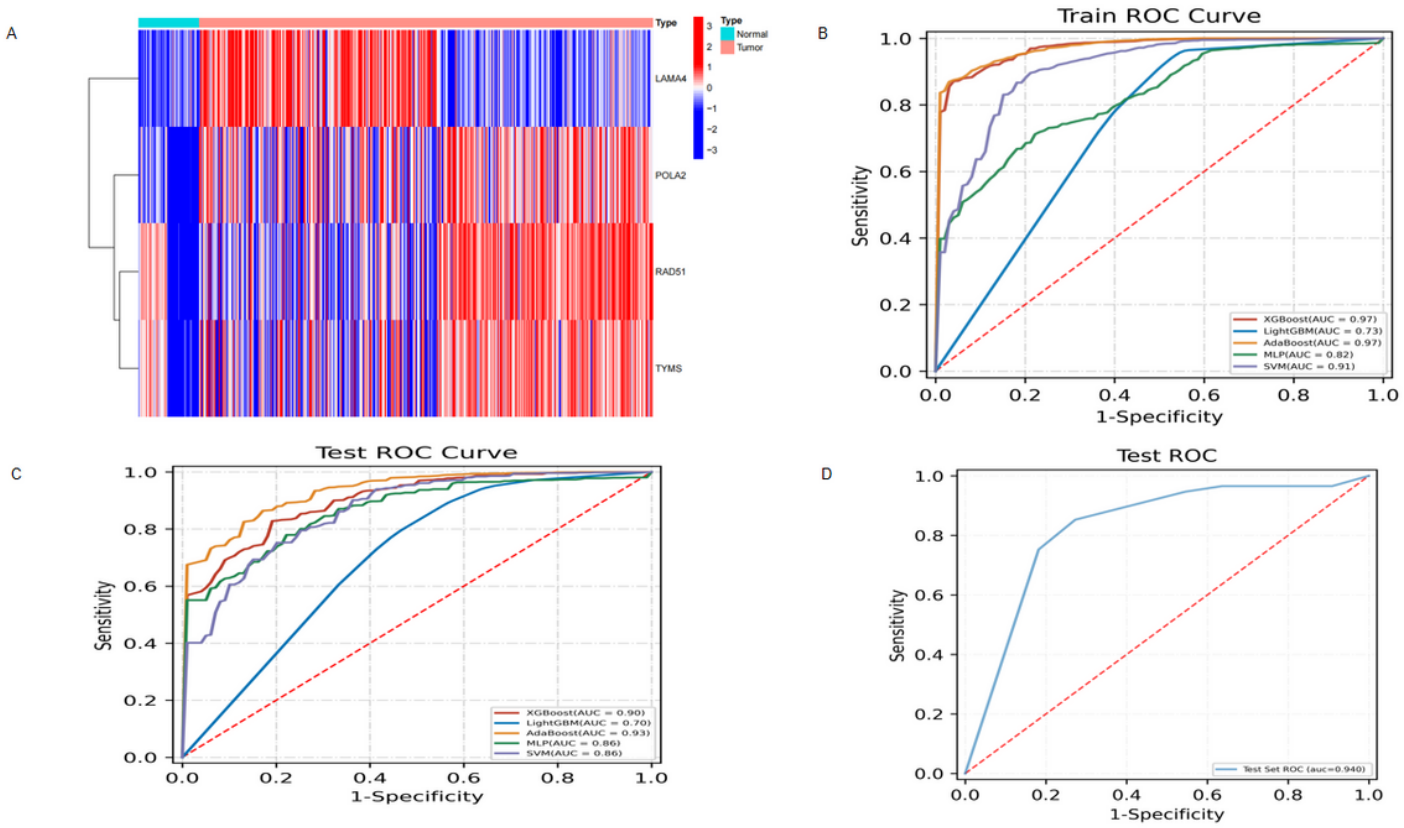

Figure 3

Establishment of model. (A) Heat map of model variables; (B) ROC curvers of classification model of HCC in training set; (B) ROC curvers of classification model of HCC in test set; (D) ROC curvers of AdaBoostClassifier model of HCC in external data test set.

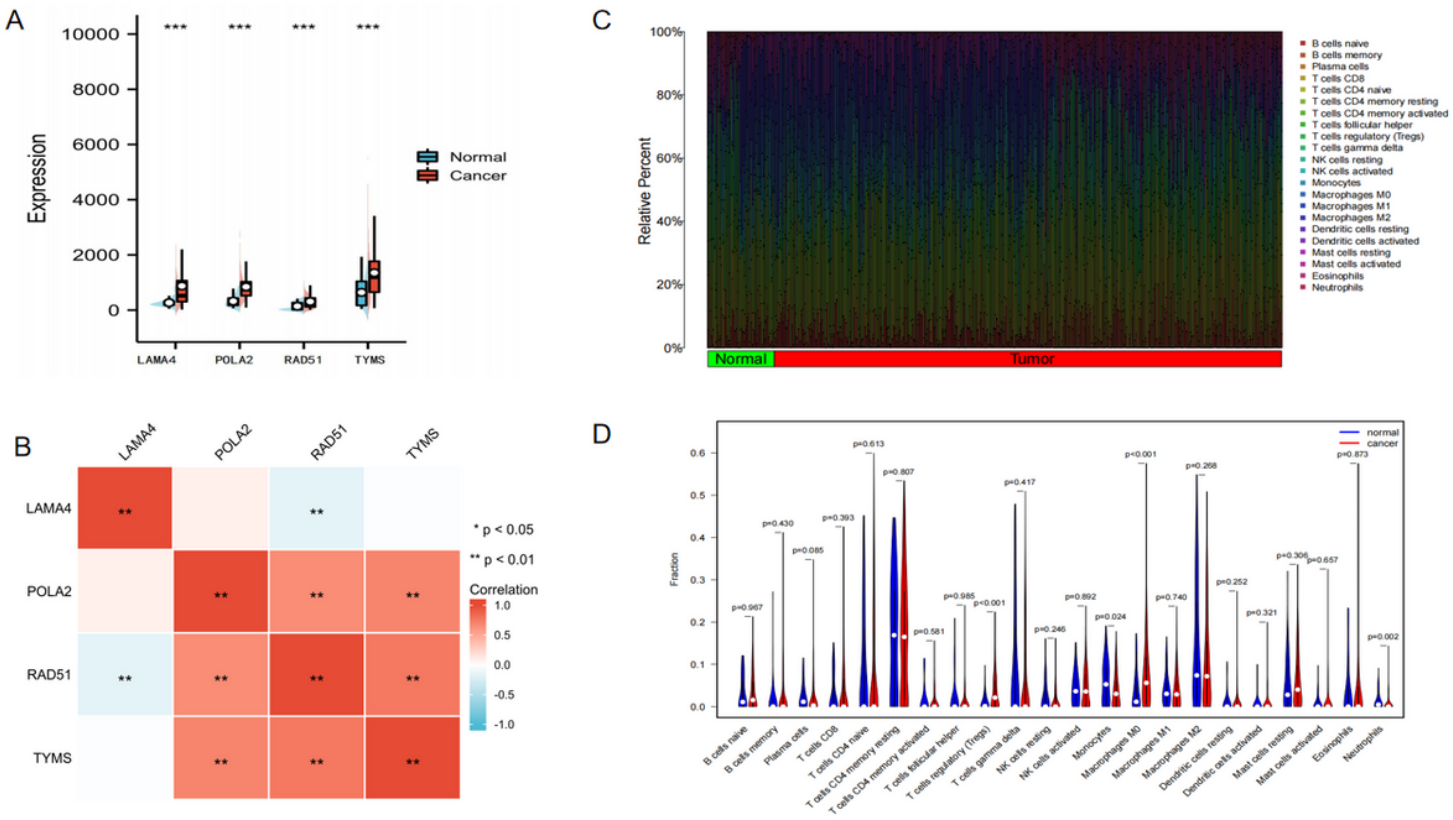

Figure 4

Bioinformatics analysis, (A) The expression of 4 grnes difference between normal and tumor tissues; (B) Graphs depicting significant associations between 4 genes; (C) Bar charts of 22 immune cell proportions in HCC and normal tissues; (D) Difference analysis of 22 immune cells infiltration between HCC and normal tissues.

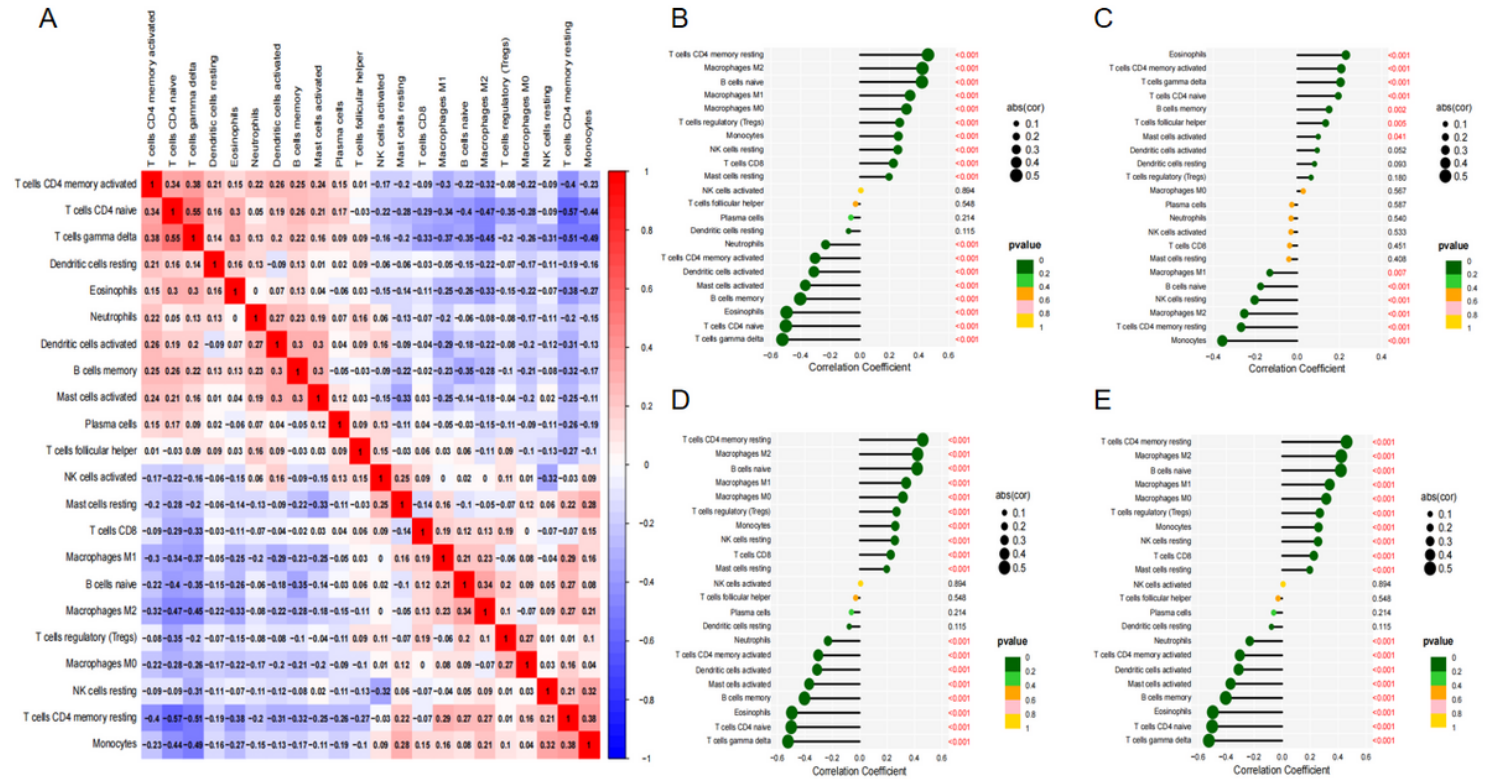

Figure 5

(A) Graphs depicting significant associations between 22 immune cells infiltration; (B) Correlation between LAMA4 and immune cells infiltration; (C) Correlation between POLA2 and immune cells infiltration; (D) Correlation between RAD51 and immune cells infiltration; (E) Correlation between TYMS and immune cells infiltration.

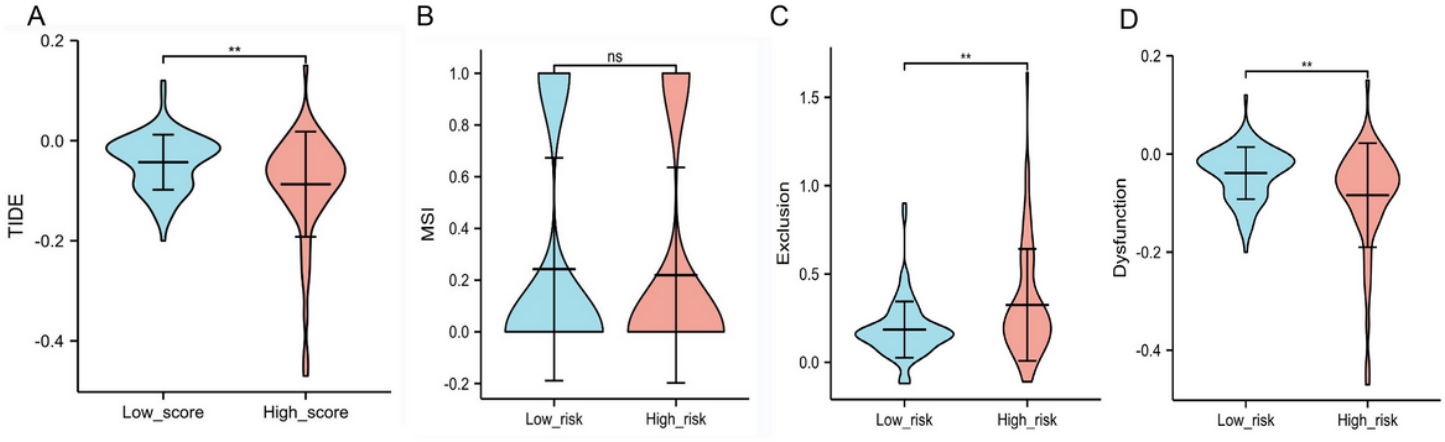

Figure 6

(A) Difference analysis of TIDE between low- and high-risk of HCC; (B) Difference analysis of MSI between low- and high-risk of HCC; (C) Difference analysis of exclusion between low- and high-risk of

HCC; (D) Difference analysis of dysfunction between low- and high-risk of HCC.

Supplementary Files

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

- supplementarytable1.txt