A four-gene-based model predicts TACE efficacy for hepatocellular carcinoma patients

Xinyi Shao
The First Affiliated Hospital of Chongqing Medical University
Tingqiao Chen
The First Affiliated Hospital of Chongqing Medical University
Yangmei Chen
The First Affiliated Hospital of Chongqing Medical University
Jiamei Wen
The First Affiliated Hospital of Chongqing Medical University
Yujie Zhang
The First Affiliated Hospital of Chongqing Medical University
Lin Liu
The First Affiliated Hospital of Chongqing Medical University
Yihuan Pu
The First Affiliated Hospital of Chongqing Medical University
Lingzhao Zhang
The First Affiliated Hospital of Chongqing Medical University
Jiayi Chen
The First Affiliated Hospital of Chongqing Medical University
Qian Li
The First Affiliated Hospital of Chongqing Medical University
Jin Chen (✉ chenjin7791@163.com)
The First Affiliated Hospital of Chongqing Medical University

Research Article

Keywords: hepatocellular carcinoma, transcatheter arterial chemoembolization, efficacy, Gene Expression Omnibus, machine learning algorithm

Posted Date: March 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1353233/v1

Abstract

Background

Primary liver cancer, of which hepatocellular carcinoma (HCC) accounts for over 90%, is the sixth most common cancer and the fifth leading cause of cancer mortality worldwide. Transcatheter arterial chemoembolization (TACE) is widely performed globally as an effective treatment for HCC patients at intermediate and advanced stages, but the efficiency varies amongst patients.

Objective

This study aimed to develop models that predict the efficacy of TACE for HCC patients.

Methods

GSE104580 from Gene Expression Omnibus was used as training and validation sets. Tumor samples were randomly allocated to training and validation sets in a ratio of 7:3. Differentially expressed genes were screened in the training sets, and efficacy-related genes were identified. Their common genes were used in LASSO L1 regression and support vector machine (SVM)-based recursive feature elimination learner to identify the best predictive four genes. After gene screening, we established the predictive model for two sets using SVMs, random forests models, and logistic regression analysis.

Result

IFIT1, LIN28B, S100A9, and SPARCL1 were significantly related to the efficacy of TACE for HCC patients. The accuracy of the training sets was 90.3%, 100%, and 91.3% using SVMs, random forests models, and logistic regression analysis, respectively. The four-gene predictive signature showed larger area under the curve values (> 80%) after receiver operating characteristic curve analysis, indicating a high predictive capacity.

Conclusion

The gene-based model constructed in this study acts as a reliable efficacy assessment tool for clinicians and will aid treatment decision-making for HCC patients.

Background

Liver cancer remains a global health challenge and ranks among the top five leading causes of cancer-related deaths worldwide[1]. Hepatocellular carcinoma (HCC) is the prominent cause of death and the

most common form of liver cancer, accounting for almost 90% of cases. Hepatitis B virus (HBV) infection is associated with the initiation and progression of HCC[2]. Hepatitis C virus (HCV) infected patients are considered to be at high risk for HCC incidence. Additionally, patients with cirrhosis are considered to be a high-risk group of HCC[3]. The current major therapeutic options include hepatectomy and liver transplantation, the mortality rate decreases, but a recurrence rate is lower at 5 years[4]. Therefore, effective and innovative treatment strategies are urgently needed.

Transcatheter arterial chemoembolization (TACE) has been the most widely used treatment for patients at intermediate and advanced stages[5]. Embolization of the hepatic arteries causes tumor necrosis. However, when the arterial blood flow is obstructed, portal blood regurgitates into the tumor through the portal vein and surrounding sinusoids, as the drainage veins of the tumor, to consequently facilitate tumor survival[6]. In the early 2000s, two randomized controlled trials reported that TACE positively influenced the aforementioned antitumor effects and survival rates compared with symptomatic treatment; however. In 101 studies involving more than 10,000 patients, the objective response rate (defined as complete or partial response) after TACE was 52.5% [95% confidence interval (CI): 43.6–61.5] [7]. However, it has the limitations of easy recurrence, poor prognosis, and incomplete effectiveness. Moreover, the survival rate after TACE treatment varies from person to person, and predicting the effectiveness of TACE treatment is still a clinical problem.

Nowadays, machine learning (ML), deep learning (DL), and artificial intelligence (AI) have undergone significant development in the medical field. ML offers various useful methodologies that can handle large dimensional datasets, efficiently and effectively evaluate a wide range of variables to construct an accurate model for prediction[8]. For example, ML developed a novel algorithm to detect non-alcoholic fatty liver disease (NAFLD)[9]. In the present work, we constructed a reliable four-gene predictive model to predict the efficacy of TACE for patients with HCC and further evaluated the accuracy and rationality of the model.

Methods

Data source

The gene expression profiles were obtained from the Gene Expression Omnibus database. First, the gene expression dataset, namely GSE104580 (total number: 147, effective treated: 81), was retrieved with Affymetrix platforms. Then background correction and normalization were performed using R software[10]. Finally, the probes were annotated using the corresponding annotation flies from the dataset.

Microarray quality control

The obtained microarray-based expression data were preprocessed in the R package. The dataset was normalized with the quantiles method to make data from all samples directly comparable (Fig. 1). The probe IDs were converted gene symbols according to the corresponding annotation files from the dataset.

For multiple probes corresponding to the same gene, the mean value of the different probes was considered as the expression value of the gene.

Random grouping method

We divided the tumor tissue microarray of HCC patients whose TACE treatment is effective or ineffective. The patients were separated into training and validation sets in a ratio of 7:3 using the stratified randomization method. The separation involved generating random values from a normal distribution with specified mean (0) and standard deviation (1) values in each gene expression series included in model construction and ordering them from high to low. The top 70 percent of patients in each gene expression series were included in the training cohort to identify and evaluate predictors—the remaining 30 percent was the validation cohort to validate the final model.

Screening for differentially expressed genes (DEGs)

Limma is a package for differential expression analysis of data arising from microarray experiments[11]. The package is designed to evaluate the differential expression of genes through linear models in the context of multi-factor experimental design. In addition, empirical Bayes and other shrinkage methods are used to assess the significance of differential gene expression.

After gene expression matrix data preprocessing, the DEGs in training sets were analyzed using the “limma” package of R software. Genes with an adjusted p value < 0.05 and absolute log2 fold change (FC) > 1.5 were identified as DEGs and selected to match the GSE104580 expression matrix for subsequent analysis.

Selection and verification of prediction-related genes

After univariate analysis by the limma package, DEGs were screened. Then, LASSO (Least Absolute Shrinkage and Selection Operator) regression was applied to select genes for predicting the efficacy of TACE in HCC using the “glmnet” package of R software[12]. The optimal values of the penalty parameter λ were determined by fivefold cross-validations to improve the reliability and objectivity of the analysis results. After obtaining the prediction-related genes calculated by the LASSO regression, the recursive feature elimination learner based on support vector machine (SVM)[13] in the e1071 package(https://cran.r-project.org/web/packages/e1071/index.html) was used to obtain the gene combination with the best prediction performance.

Building and validating a predictive model

The predictive genes were screened by Lasso Regression and SVM-based recursive feature elimination learner. RandomForests package[14], e1071 package, and glm function were used to construct random forest, support vector machine, and logistic regression models on the training data set. The models were compared in terms of area under the receiver operating characteristic (ROC) curve (AUC), accuracy, specificity, and sensitivity. The validation sets were used to confirm the predictive ability of the model.

Results

Identification of the TACE-associated DEGs in HCC

A flow chart of the analysis procedure shown in Fig. 2 describes our study more clearly. First, the gene expression matrix of HCC patients whose TACE treatment is effective or ineffective was analyzed using the limma package. A total of 66 significant DEGs were identified (28 upregulated and 38 downregulated; Fig. 3A), of which 14 DEGs had |logFC|>2 (Fig. 3B). The Euclidean distance was calculated between any two patients in the discovery cohort and condensed into two-dimensional points using a nonlinear dimensionality reduction algorithm t-SNE(Fig. 3C), and subsequently, two patient clusters were determined by which is positive or negative. The top 15 genes with the highest value of |logFC| was showed in Fig. 3D. Then, 66 DEGs were used for hierarchical cluster analysis for patients with TACE treatment.

Identification gene predictive makers

The optimal values of the penalty parameter λ from minimum partial-likelihood bias were determined by fivefold cross-validations (Fig. 4A-B). Fifteen genes were selected as the best predictors. Interestingly, only four genes achieved the selected criteria (|logFC| >2) in the best predictive gene from the LASSO Regression: HK2, MEP1A, ADH1B, LIN28B, and coefficient of SPARCL1 were the highest (Fig. 4C). All four predictive genes presented statistically significant differences in the training dataset between HCC patients whose TACE treatment is effective or ineffective (Fig. 5C-F), the result of unsupervised hierarchical clustering of four genes based on Euclidean distance grouped all samples into two clusters, and the majority of effective treatment samples clustered together in cluster II, accounting for 77.1% (Fig. 5A). LogFC of four genes secondary selected by LASSO regression, and logFC of SPARCL1 were the highest (Fig. 5B). SVM-based recursive feature elimination learner was used to do the secondary feature selection. A combination of four genes was selected as the best predictor of efficacy: IFIT1, LIN28B, S100A9, and SPARCL1.

Construction of prediction model.

The prediction model was constructed using four predictive genes selected through lasso regression and SVM. In the training sets, the predictive accuracy of the three models was 90.3%, 100%, 91.3% (Table 1), the AUCs of the three models were 86.0%, 100%, 88.0%. The predictive accuracy of the three models in the validation sets was above 86% (Table 2, Fig. 4D). In the validation sets, the AUCs of the three models were 82%, 85.3%, 82.5% (Fig. 6), the specificity and sensitivity of SVM were 78.9% and 91.7%, the specificity and sensitivity of random forests models were 73.7%, and 95.8%, the specificity, and sensitivity of logistic regression analysis were 78.9% and 91.7% (Table 2, Fig. 6). The results suggested that the model had strong prediction ability in the training and the validation sets and can significantly distinguish the efficacy of TACE in HCC patients. Because SVM and RF are both non-parametric models, they cannot evaluate the weight of each gene in the construction of the model. However, because the logistic regression model is parametric, the weight of four genes can be calculated in this model. The

partial regression coefficients of LIN28B, S100A9, IFIT1, and SPARCL1 were − 3.897, -3.374, 3.950, 4.833 (Table 3). In addition, the ROC curves were plotted to assess the prediction accuracy in the gene expression matrix of 147 patients. The AUCs of LIN28B, S100A9, IFIT1, SPARCL1 were 72.9%, 72.7%, 76.9%, 84.2%, respectively (Fig. 6). Because SPARCL1 had a large weight in the logistic regression model, and the predicted AUC value was significantly higher than the other genes. We further identified optimal cut-off values of SPARCL1 expression as 10.64 through the Youden index. Using optimal cut-off values to define the effectiveness of TACE, the accuracy, specificity, and sensitivity of SPARCL1 were 76.9%, 92.4%, 64.2% (Table 4).

Table 1
The prediction ability of three models in the training sets

	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	AUC
SVM	0.903	0.830	0.953	0.936	0.877	0.86
Random forests	1	1	1	1	1	1
Logistic regression	0.913	0.842	0.960	0.936	0.894	0.88

Table 2
The prediction ability of three models in the validation sets

	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	AUC
SVM	0.860	0.789	0.917	0.882	0.846	0.82
Random forests	0.860	0.736	0.958	0.933	0.821	0.853
Logistic regression	0.860	0.789	0.917	0.882	0.846	0.825

Table 3
partial regression coefficiens of genes in the model of logistic regression

Offset value	LIN28B	S100A9	IFIT1	SPARCL1
1.419626	-3.89717	-3.37414	3.949936	4.833497

Table 4
The prediction ability of SPARCL1

	accuracy	sensitivity	specificity	Positive predictive value	Negative predictive value	AUC
SPARCL1	0.769	0.924	0.642	0.678	0.912	0.842

Discussion

In the present study, three predictive models: SVM, random forest, and logistic regression, were constructed based on a four-gene signature to predict effective and ineffective TACE in HCC patients. The model based on the random forest algorithm had the best AUC value amongst the three models with AUC values of 100% and 85.2% in the training and testing datasets, respectively. All models showed good performance in the training and testing datasets with high accuracy and AUC values. The four-gene signature was identified by combining traditional multiple bioinformatic analysis and feature selection methods in a machine learning algorithm. Interestingly, the expression of all four genes was highly correlated with the prognosis of HCC patients (Fig. 7); in general, significantly higher expressions of LIN28B and S100A9 were related to a gradually worse prognosis of HCC. However, the significantly higher expression of IFIT1 and SPARCL1 were related to a better prognosis of HCC. It is noteworthy that the most predictive single gene for effective and ineffective TACE in HCC patients is SPARCL1, with an AUC value of 84.2% (Fig. 5F).

During the last two decades, the incidence of HCC significantly increased, but the mortality was not decreased[15]. TACE is widely performed globally as an effective treatment for inoperable HCC, but its efficacy varies greatly. At present, the six-and-twelve score is used to predict the prognosis of stratifying recommended TACE candidates[16]. However, the risk of selection bias is unavoidable in observational studies. Therefore, predictive biomarkers are urgently needed to predict the effectiveness of TACE and to outline an individualized treatment plan for HCC patients.

In this study, four genes were identified for constructing the predictive model. These genes were found to be involved in several cancer-related activities. S100 A9 belongs to the S100 family, mainly expressed in neutrophils and monocytes, and plays an important role in regulating inflammation and innate immunity[17]. Furthermore, S100A9 is up-regulated in several solid tumors, including colorectal, prostate, breast, and liver cancers[18]. Importantly, overexpression of S100A9 is positively correlated with poor differentiation, tumor invasion, metastasis, and poor clinical outcomes in these cancers, indicating its key role in mediating tumor progression.

SPARCL1 is a potential tumor suppressor gene in most tumors[19]. The downregulation of SPARCL1 is considered to be regulated by epigenetic modifications (including DNA methylation). In addition, SPARCL1 regulates cell viability, migration, invasion, cell adhesion, and drug resistance. The downregulation of SPARCL1 is associated with increased mortality in patients with liver cancer[20]. However, the related mechanism between SPARCL1 and HCC is still unclear.

The IFIT gene family consists of four genes. They perform various cellular functions by mediating protein-protein interaction and forming multi-protein complexes with cells and viral proteins through different TPR motifs[21]. IFIT protein is involved in cancer progression and metastasis. The expression of IFIT family members (IFIT1, IFIT2, IFIT3) decreased in HCC[22], IFIT1, IFIT2, and IFIT3 may be involved in the progress of HCC. IFIT1 or IFIT3 silencing can reduce the expression of IL-17 and IL-1β and reduce the migration ability of HCC cells[23].

Lin28 is a major regulator of the microRNAs let-7 family[24]. Lin28B is overexpressed in human hepatoma cells and clinical samples[25, 26] by promoting malignant transformation[30, 31], promoting tumor-associated inflammation[27, 28], reprogramming metabolism, obtaining immortality, and avoiding immune destruction[29]. Cheng et al. reported that Lin28B was associated with high tumor grade, large tumor volume, AJCC stage, clinical liver cancer stage, and recurrence in Barcelona. In addition, clinical, epidemiological studies have shown that Lin28B is related to HCC susceptibility and to the overall survival rate of patients[30, 31, 32].

However, several limitations of the current study should be considered. Firstly, our study only focused on the samples from the GSE104580 dataset. The numbers of patients in the Gene Expression Omnibus (GEO) database are relatively small. More patients and clinical information should be collected to validate the stability of the model further. Secondly, some genes might be excluded because of our rigorous screening criteria. Thirdly, our study provides evidence that four novel genes are significantly related to the survival of HCC patients; more experiments will be needed for validation or even correction and confirm the KEGG pathway analysis and GO enrichment results.

In conclusion, we constructed a four-gene predictive model by performing logistic regression analysis and 5-fold cross-validation based on datasets from GEO. The stability and accuracy were further assessed in three independent models. The proposed algorithm obtains stable results with high accuracy and low bias, superior prediction performance is achieved. Future studies suggested that genes from the predictive model are involved in several cancer-related biological processes. This predictive model has provided new insight into the prediction of the effectiveness of TACE and has potential prognostic and therapeutic implications for HCC.

To our best knowledge, this study is the first to use machine learning techniques to predict the effectiveness of TACE in patients with HCC through genes expression. TACE is currently a preferred surgical approach for patients with advanced HCC, but its treatment efficacy varies greatly, and the five-year survival rate is still low. Therefore, it is crucial to predict the therapeutic effectiveness of TACE before operation and establish personalized treatment strategies to manage the prognosis of HCC patients. Thus, our predictive model based on a four-gene signature may have good prospects in clinical practice.

Conclusions

In conclusion, we developed 4-gene predictive models, and all models had high sensitivity and specificity to predict the efficacy of TACE for HCC patients. In addition, the expression of individual genes shows predictive power in efficacy assessment.

Abbreviations

HCC hepatocellular carcinoma; TACE Transcatheter arterial chemoembolization; SVM support vector machine; ML machine learning; DL deep learning; AI artificial intelligence; DEGs differentially expressed

genes; ROC receiver operating characteristic; RF random forest; HBV hepatitis B virus; HCV hepatitis C virus; LASSO Least Absolute Shrinkage and Selection Operator.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed in the current study are available in the GEO (https://www. ncbi.nlm.nih.gov/geo/).

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by National Natural Science Foundation of China (n82073462 and n81773307) and Chongqing science and technology commission (cstc2018jcyjAX0195).

Authors' Contributions

JC, XS and TC wrote the manuscript. JC and YC designed the original research. TC, XS, YZ, JC, QL analyzed and interpreted the data. JW, LL, YP and LZ collected and preprocessed the data. JC and XS are in charge of the whole research conduction and paper writing. All of the authors reviewed the manuscript before submission and approved the final manuscript.

Acknowledgments

Not applicable.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69–90. doi: 10.3322/caac.20107.
2. Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, Allen C, et al. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. JAMA Oncol. 2017;3(12):1683–91. doi: 10.1001/jamaoncol.2017.3055.

3. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology. 2018;67(1):123–33. doi: 10.1002/hep.29466.
4. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. J Hepatol. 2018;69(1):182–236. doi: 10.1016/j.jhep.2018.03.019.
5. Llovet JM, Bruix J. Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival. Hepatology. 2003;37(2):429–42. doi: 10.1053/jhep.2003.50047.
6. Ekelund L, Lin G, Jeppsson B. Blood supply of experimental liver tumors after intraarterial embolization with gelfoam powder and absolute ethanol. Cardiovasc Intervent Radiol. 1984;7(5):234–9. doi: 10.1007/BF02553141.
7. Cammà C, Schepis F, Orlando A, Albanese M, Shahied L, Trevisani F, et al. Transarterial chemoembolization for unresectable hepatocellular carcinoma: meta-analysis of randomized controlled trials. Radiology. 2002;224(1):47–54. doi: 10.1148/radiol.2241011262.
8. Badrick T, Banfi G, Bietenbeck A, Cervinski MA, Loh TP, Sikaris K. Machine Learning for Clinical Chemists. Clin Chem. 2019;65(11):1350–6. doi: 10.1373/clinchem.2019.307512.
9. Yip TC, Ma AJ, Wong VW, Tse YK, Chan HL, Yuen PC, et al. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. Aliment Pharmacol Ther. 2017;46(4):447–56. doi: 10.1111/apt.14172.
10. Zhang Y, Szustakowski J, Schinke M. Bioinformatics analysis of microarray data. Methods Mol Biol. 2009;573:259–84. doi: 10.1007/978-1-60761-247-6_15
11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. doi: 10.1093/nar/gkv007.
12. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.
13. Meyer D, Dimitriadou, E, Hornik, K, Weingessel A & Leisch F. R Package Version 1.6–8; TU Wien:. e1071. 2017.
14. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. 2003;43(6):1947–58. doi: 10.1021/ci034160g.
15. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. Nat Rev Dis Primers. 2021;7(1):6. doi: 10.1038/s41572-021-00245-6.
16. McGlynn KA, Petrick JL, London WT. Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and regional variability. Clin Liver Dis. 2015;19(2):223–38. doi: 10.1016/j.cld.2015.01.001.
17. Duan L, Wu R, Zhang X, Wang D, You Y, Zhang Y, et al. HBx-induced S100A9 in NF-κB dependent manner promotes growth and metastasis of hepatocellular carcinoma cells. Cell Death Dis.

2018;9(6):629. doi: 10.1038/s41419-018-0512-2.

18. Srikrishna G, Freeze HH. Endogenous damage-associated molecular pattern molecules at the crossroads of inflammation and cancer. Neoplasia. 2009;11(7):615–28. doi: 10.1593/neo.09284.
19. Yan Q, Sage EH. SPARC, a matricellular glycoprotein with important biological functions. J Histochem Cytochem. 1999;47(12):1495–506. doi: 10.1177/002215549904701201.
20. 19. Lau CP, Poon RT, Cheung ST, Yu WC, Fan ST. SPARC and Hevin expression correlate with tumour angiogenesis in hepatocellular carcinoma. J Pathol. 2006;210(4):459–68. doi: 10.1002/path.2068.
21. Pidugu VK, Pidugu HB, Wu MM, Liu CJ, Lee TC. Emerging Functions of Human IFIT Proteins in Cancer. Front Mol Biosci. 2019;6:148. doi: 10.3389/fmolb.2019.00148.
22. Yang Y, Zhou Y, Hou J, Bai C, Li Z, Fan J, et al. Hepatic IFIT3 predicts interferon-α therapeutic response in patients of hepatocellular carcinoma. Hepatology. 2017;66(1):152–66. doi: 10.1002/hep.29156.
23. Liu G, Sun J, Yang ZF, Zhou C, Zhou PY, Guan RY, et al. Cancer-associated fibroblast-derived CXCL11 modulates hepatocellular carcinoma cell migration and tumor metastasis through the circUBAP2/miR-4756/IFIT1/3 axis. Cell Death Dis. 2021;12(3):260. doi: 10.1038/s41419-021-03545-7.
24. Panella M, Mosca N, Di Palo A, Potenza N, Russo A. Mutual suppression of miR-125a and Lin28b in human hepatocellular carcinoma cells. Biochem Biophys Res Commun. 2018;500(3):824–7. doi: 10.1016/j.bbrc.2018.04.167.
25. Tian N, Shangguan W, Zhou Z, Yao Y, Fan C, Cai L. Lin28b is involved in curcumin-reversed paclitaxel chemoresistance and associated with poor prognosis in hepatocellular carcinoma. J Cancer. 2019;10(24):6074–87. doi: 10.7150/jca.33421.
26. Zhang J, Hu K, Yang YQ, Wang Y, Zheng YF, Jin Y, et al. LIN28B-AS1-IGF2BP1 binding promotes hepatocellular carcinoma cell progression. Cell Death Dis. 2020;11(9):741. doi: 10.1038/s41419-020-02967-z.
27. Kugel S, Sebastián C, Fitamant J, Ross KN, Saha SK, Jain E, et al. SIRT6 Suppresses Pancreatic Cancer through Control of Lin28b. Cell. 2016;165(6):1401–15. doi: 10.1016/j.cell.2016.04.033.
28. Iliopoulos D, Hirsch HA, Struhl K. An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. Cell. 2009;139(4):693–706. doi: 10.1016/j.cell.2009.10.014.
29. Ma L, Zhao Q, Chen W, Zhang Y. Oncogene Lin28B increases chemosensitivity of colon cancer cells in a let-7-independent manner. Oncol Lett. 2018;15(5):6975–81. doi: 10.3892/ol.2018.8250.
30. Permuth-Wey J, Kim D, Tsai YY, Lin HY, Chen YA, Barnholtz-Sloan J, et al. LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. Cancer Res. 2011;71(11):3896–903. doi: 10.1158/0008-5472.CAN-10-4167.
31. Diskin SJ, Capasso M, Schnepp RW, Cole KA, Attiyeh EF, Hou C, et al. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. Nat Genet. 2012;44(10):1126–30. doi: 10.1038/ng.2387.

32. Chen AX, Yu KD, Fan L, Li JY, Yang C, Huang AJ, et al. Germline genetic variants disturbing the Let-7/LIN28 double-negative feedback loop alter breast cancer susceptibility. PLoS Genet. 2011;7(9):e1002259. doi: 10.1371/journal.pgen.1002259.

Figures

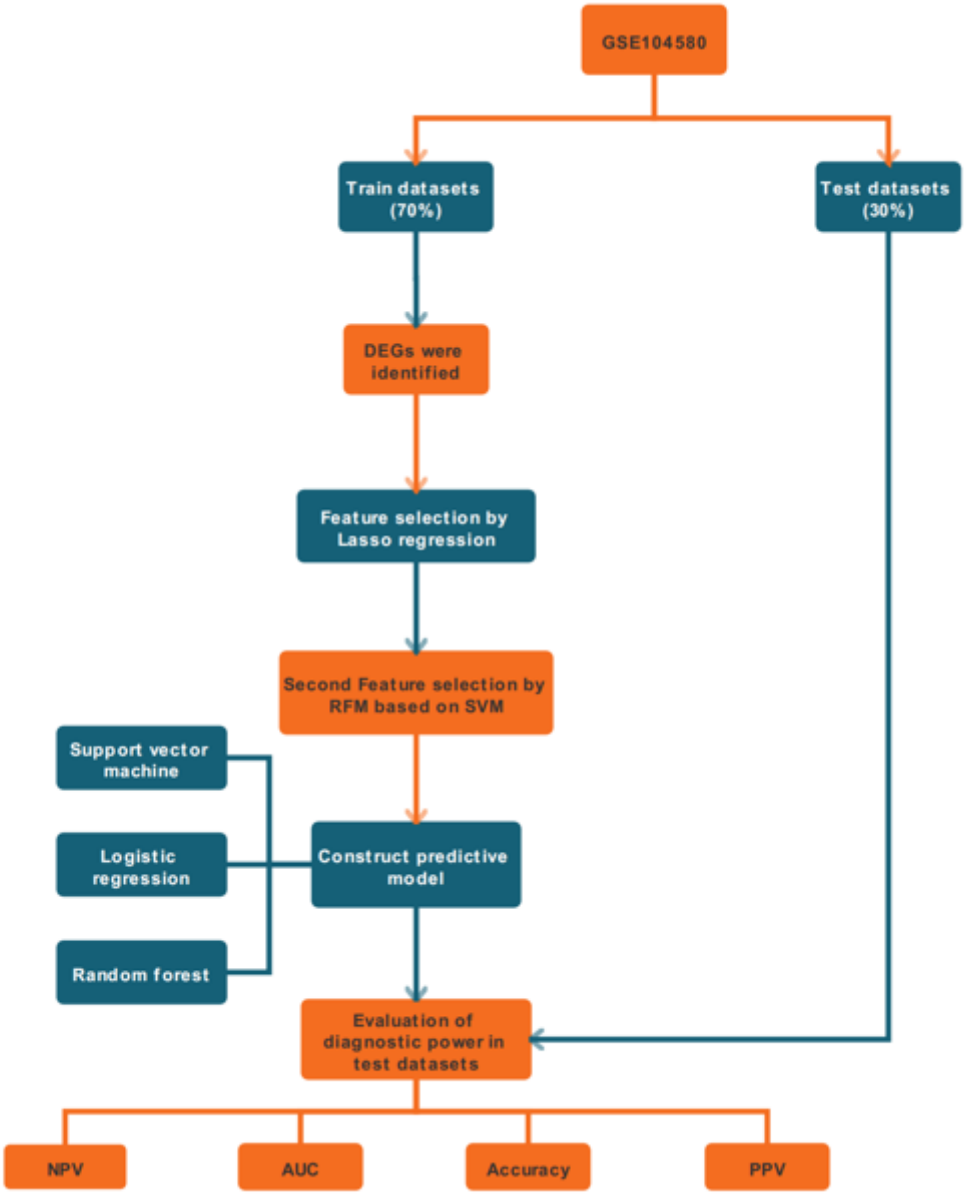

Figure 1

Overall workflow describing the process used to develop and validate the prognostic model.

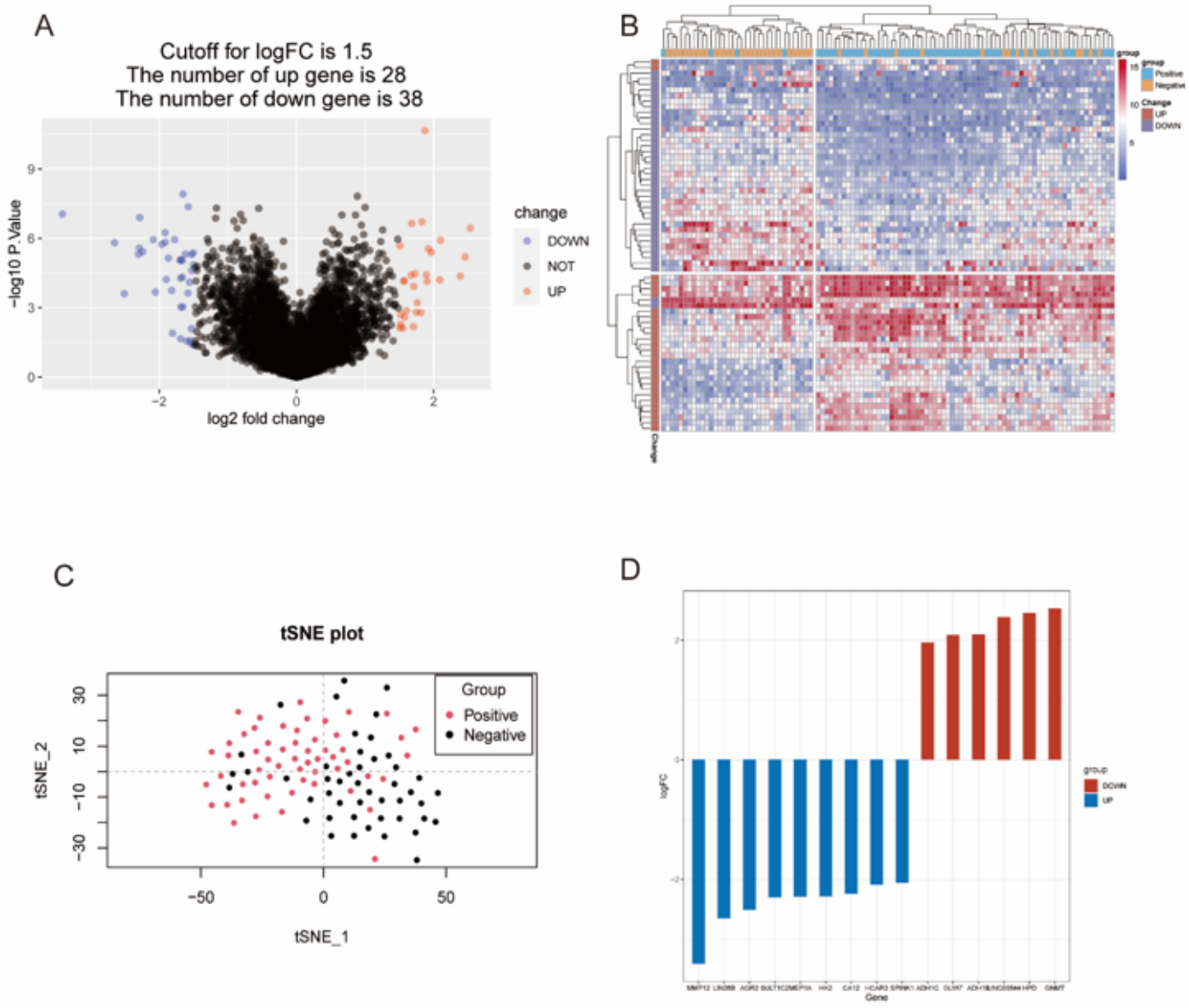

Figure 2

Data processing, screening of the DEGs. A Heatmap of top 66 TACE-associated DEGs in HCC; B Volcano plot of differentially expressed genes in HCC patients whose TACE treatment is effective compared with when treatment is ineffective. Red nodes represent the significantly up-regulated genes with logFC=2 and p<0.05. Purple nodes represent the significantly down-regulated genes with logFC=-2 and p<0.05; C Dimension reduction by t-SNE algorithm; D LogFC of genes first selected by Lasso regression.

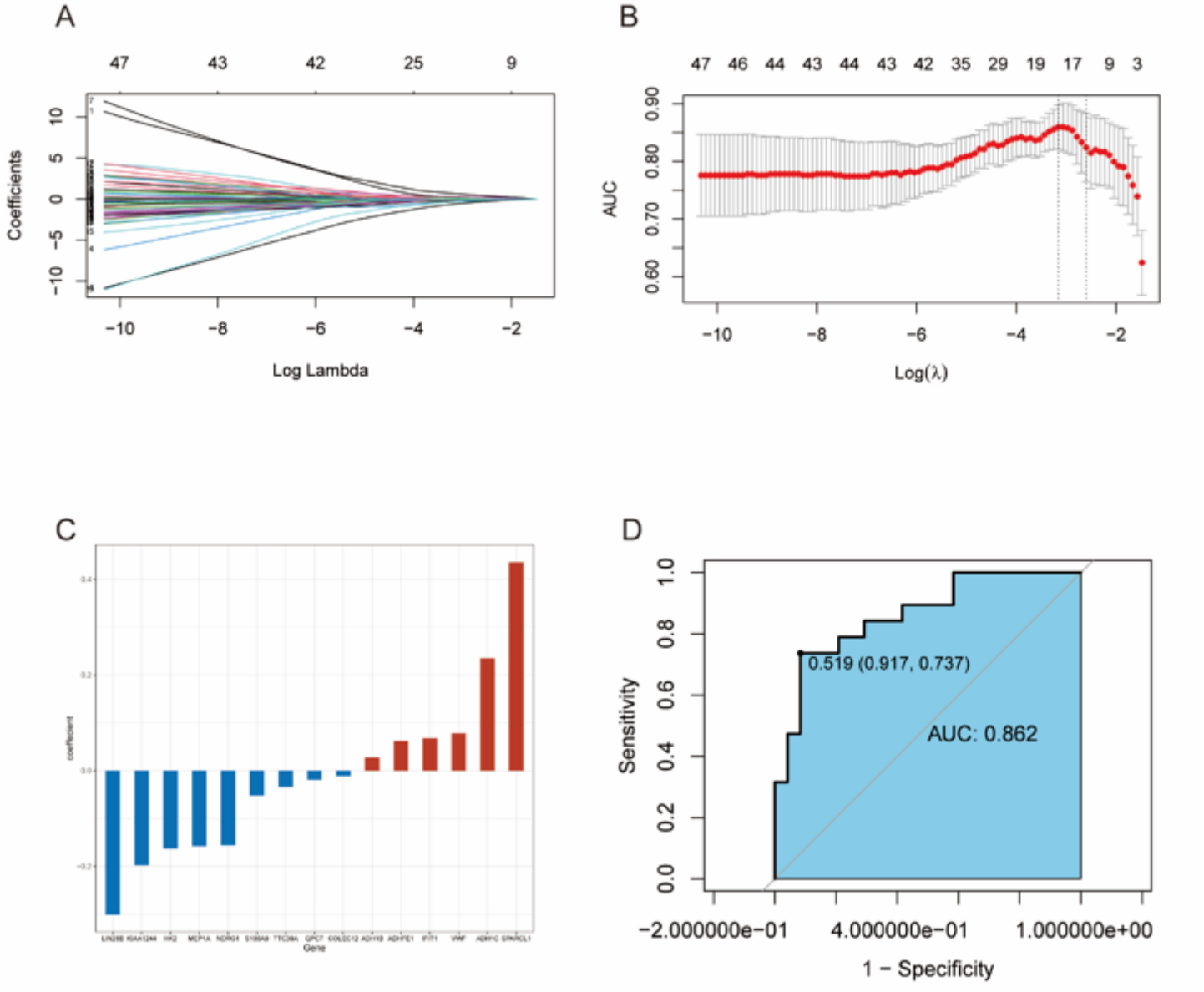

Figure 3

Selection of predict-related genes for the efficacy of TACE. A LASSO coefficient profile of the genes associated with the efficacy of TACE of HCC. B five-fold cross-validation for parameter selection. The AUC is a plot against log(λ), which is a tuning parameter, and dotted lines are drawn at optimal AUC values by minimum criteria and 1-se criteria; C Coefficient of genes selected by Lasso regression; D receiver-operating characteristic (ROC) curve in the test dataset, the area under curve (AUC) is 86.2% for differentially predicting from the efficacy of TACE.

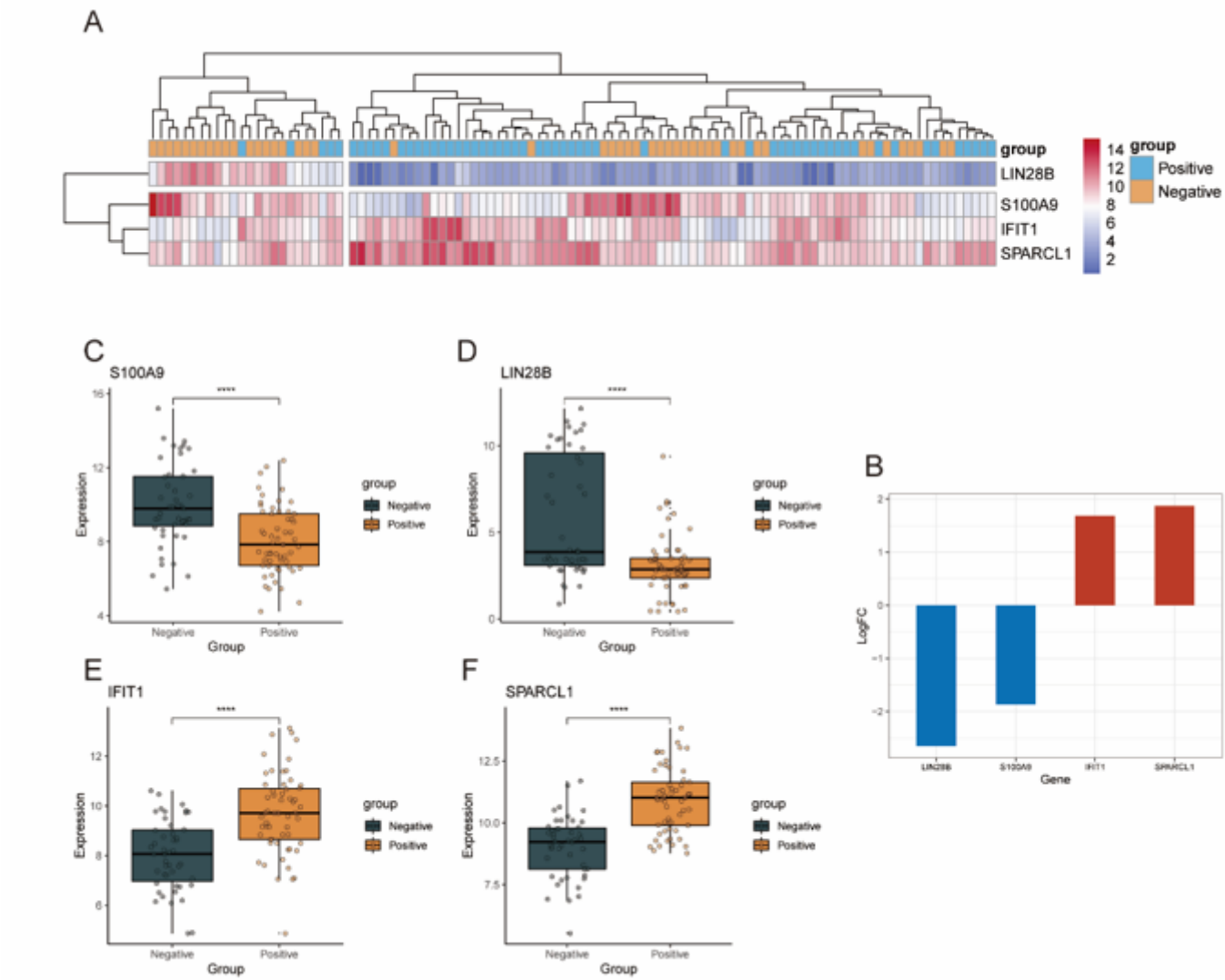

Figure 4

Expression prediction-related genes for the efficacy of TACE in the training dataset. A Heatmap of genes expression. B LogFC of genes selected by Lasso regression. C-F Statistical differences of genes between positive and negative outcomes in training data.

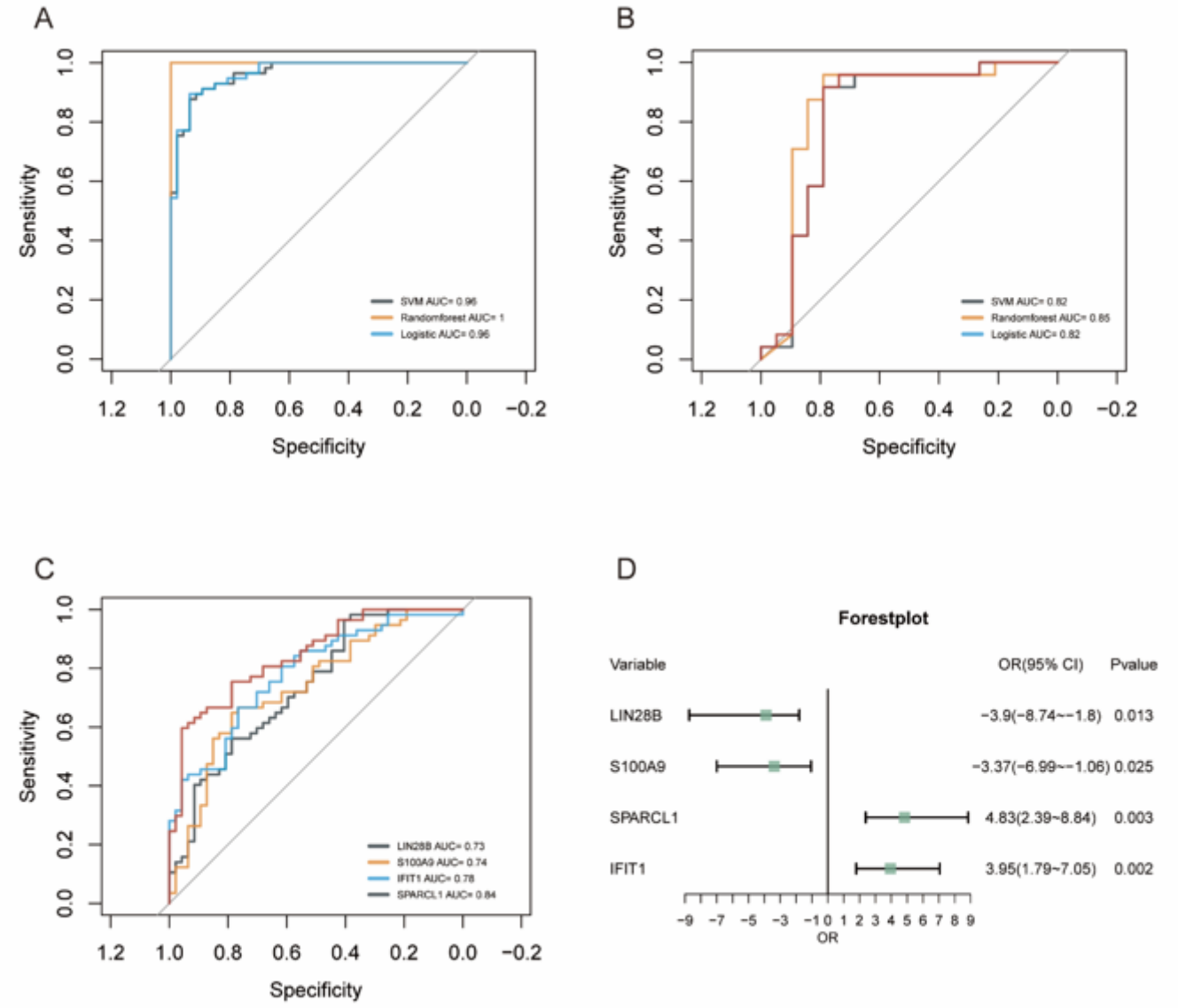

Figure 5

Evaluation of models constructed by machining learning algorithm and traditional methods. A ROC analysis of train dataset for predicting the efficacy of TACE by predictive models. B ROC analysis of test dataset for predicting the efficacy of TACE by predictive models. C ROC analysis of training dataset for predicting the efficacy of TACE by prediction-related genes. D ROC analysis of test dataset for predicting the efficacy of TACE by four prediction-related genes. E Forest plot display the coefficient of four genes in logistic regression.

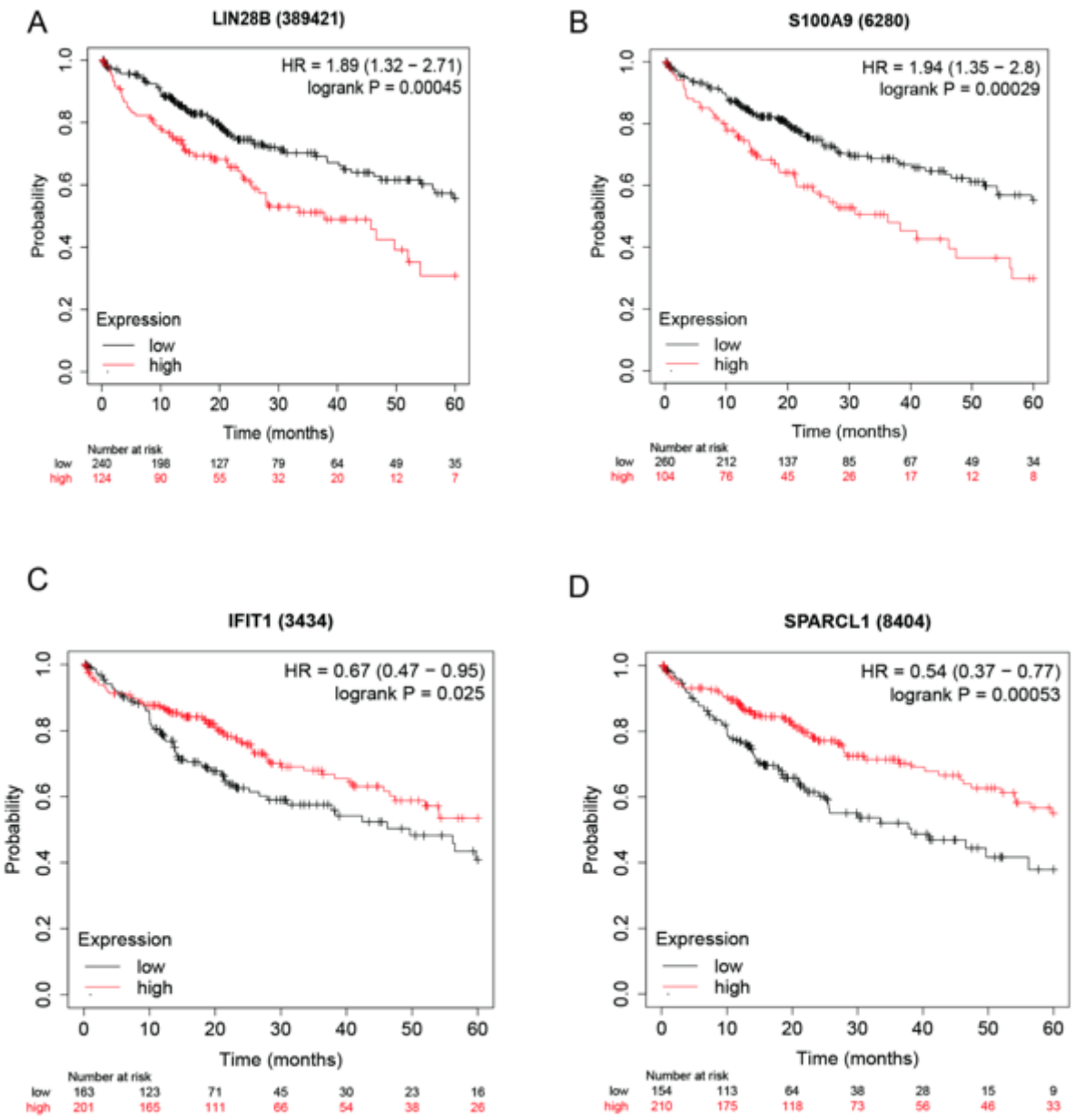

Figure 6

Construction and validation of prediction-related gene signature using GEO dataset. A-D Survival curve of low- and high-risk groups stratified by four prediction-related genes.

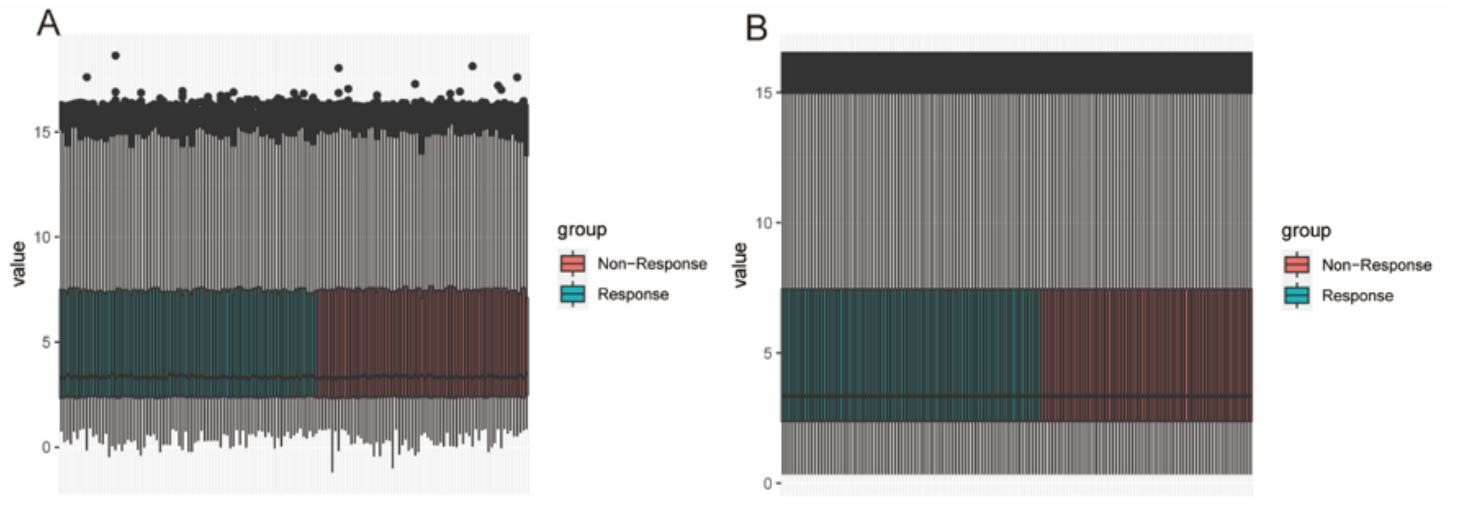

Figure 7

Data preprocesses. A Before GSE104580 quantile normalized; B After GSE104580 quantile normalized.