

Prognostic value of a metastasis-related gene signature in hepatocellular carcinoma

Zhou Zhou

Leids Universitair Medisch Centrum

Yinan Jiang (✉ jiangyn@whu.edu.cn)

Wuhan University Zhongnan Hospital <https://orcid.org/0000-0002-3602-7278>

Primary research

Keywords: Hepatocellular carcinoma, WGCNA, LASSO, Prognosis, Metastasis

Posted Date: February 17th, 2020

DOI: <https://doi.org/10.21203/rs.2.23617/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Surgery is a potential curative treatment for hepatocellular carcinoma (HCC), but postoperative recurrence occurs in majority of patients. Currently, there are no robust biomarkers to predict the disease progression or recurrence.

Methods: Weighted gene correlation network analysis (WGCNA) was used to construct co-expression network. The least absolute shrinkage and selection operator (LASSO) method was applied to develop prognostic model. Survival analysis, gene set variation analysis (GSVA) and gene set enrichment analysis (GSEA) were performed to assess the performance.

Results: By using GSE14520 dataset, 16 hub genes associated with HCC metastasis were screened. A signature of 5 metastasis-related genes was subsequently constructed by incorporating TCGA LIHC dataset. HCC patients with high risk score have low survival rate, low disease free survival rate and high recurrence rate. The prognostic value of this 5-mRNA signature was further verified in pan-cancer datasets. A nomogram was built with excellent performance and potential clinical application for prognosis. GSVA and GSVA revealed that metabolic pathways are distinct between high- and low-risk groups.

Conclusions: We have established a 5-mRNA signature and a nomogram that can efficiently predict metastasis, recurrence and patient survival in HCC.

Background

Hepatocellular carcinoma (HCC) is currently the fourth leading cause of cancer death worldwide [1]. Although partial hepatectomy is the recommended first-line treatment for primary HCC, 70% of patients have tumor recurrence and 80% cancer-related mortality occurs within 5 years, which is largely attributed to metastasis [2, 3, 4]. Unlike other carcinomas, HCC occasionally metastasizes to distant organs, whereas frequently metastasizes within the liver and liver vasculature. Recurrence by metastasis seems to occur mainly within the first two years after resection [5, 6]. However, the current known clinic-pathological risk factors are not sufficient to recognize HCC patients with a high risk of tumor progression. In fact, none of the molecular classifications of HCC that have been proposed so far can decently predict disease progression or recurrence [7, 8].

Numerous studies have attempted to develop molecular biomarkers for HCC prognosis. Serum alpha-fetoprotein (AFP) is the most widely used biomarker for HCC diagnosis, and has been explored for prognosis. However, there is increasing debate regarding its actual value for HCC surveillance [9]. Other molecules, such as AFP-L3, Des-γ-carboxyprothrombin, SALL4, CDK1 and Sox12, have also been investigated for the early diagnosis and surveillance of HCC patients [10, 11]. Given the heterogeneous nature of both the patients and tumors, the performance of a single biomarker-based approach is always limited [12]. Thus, the combination of multiple biomarkers is necessary to maximize the diagnostic and predictive capability. In this respect, a seven-microRNA expression signature and a four-gene-based model

have been proposed to predict HCC disease progression or patient survival [9, 13]. Their potential for clinical implementation however remains to be further validated.

Recently, the development of high-throughput profiling, sequencing technology and bioinformatics has provided new ways for cancer research. By utilizing the publically available datasets and advanced bioinformatics tools, we aim to systematically explore biomarkers for HCC prognosis. The weighted gene correlation network analysis (WGCNA) provides R functions for weighted correlation network analysis and can be used as a data exploratory tool or a gene screening (ranking) method to find clusters (modules) of highly correlated genes [14]. It has been widely used to find hub genes in various cancers, such as breast cancer, adrenocortical carcinoma, lung carcinoma and pancreatic cancer [15, 16, 17, 18]. The least absolute shrinkage and selection operator (LASSO) method was originally designed for regression analysis [19]. Lately, it has been applied to many fields, including the construction of prognostic models for various cancers [20, 21, 22, 23].

Given the importance of metastasis in determining the outcome of HCC patients after surgical operation, in this study, we firstly used WGCNA to identify hub genes associated with HCC metastasis. By using LASSO method, we then constructed a 5-mRNA signature and extensively tested for predicting patient survival. Next, we built a nomogram based on the 5-mRNA signature combined with other clinical factors. We have confirmed the clinical utility by decision curve analysis.

Materials And Methods

Data collection

We searched on Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) with keywords “HCC”, chose study type as “expression profiling by array”, top organisms as “Homo sapiens” and sorted by “number of samples (high to low)”. Meanwhile, we downloaded RNA-sequencing data of HCC from The Cancer Genome Atlas (TCGA) database (<https://genome-cancer.ucsc.edu/>).

WGCNA analysis

Firstly, we performed data preprocessing to remove outlier samples. The “WGCNA” R package was then used to construct the co-expression network. We identified a set of genes based on the module connectivity, measured by absolute value of the Pearson's correlation ($\text{cor.geneModuleMembership} > 0.8$) and clinical trait relationship, measured by absolute value of the Pearson's correlation ($\text{cor.geneTraitSignificance} > 0.2$). The common genes were selected to perform further study.

Construction of a prognostic mRNA classifier

Genes from the module most related to tumor metastasis risk were used. Based on “glmnet” R package, Lasso cox regression analysis was applied to build an optimal prognostic signature of HCC samples by using TCGA LIHC. Time-dependent receiver operating characteristic curve (ROC) analysis by using

“survivalROC” R package was used to calculate the area under curve (AUC) and to evaluate the prediction accuracy of the model.

Analysis of the relation of the prognostic model with clinical characteristics

To determine whether the prognostic model was significant in combination with other clinical characteristics, all the clinical variables in TCGA, including stage, grade, N stage, M stage, age and gender, were divided into two groups. Patients were stratified into stage I/II and stage III/IV subgroups, grade I/II and grade III/IV subgroup, N0 and N1/Nx subgroup, M0 and M1/Mx subgroup, age < 65y and age \geq 65y subgroup, male and female subgroup. Overall survival analyses were performed in each subgroup.

Building and validating a predictive nomogram

The “regplot” R package was used to plan the nomogram. Calibration plots were generated to evaluate whether the predicted result (by nomogram) was close to the actual outcome. The x-axis represents nomogram predicted probability of survival and the y-axis represents the actual survival. The 45-degree line represents the perfect nomogram performance. Decision curve analysis (DCA) was used to determine whether the predictive model is clinically useful.

Pan-cancer outcome validation

To further investigate the clinical value of the signature, pan-cancer validation of overall survival was performed based on KM plotter database (<http://kmplot.com>). 19 types of tumors were included (n = 7008). Since the signature has high correlation with tumor metastasis, gene expression profile of human breast cancer and prostate cancer with metastasis information was downloaded and analyzed.

Functional enrichment analysis

The Gene Set Variation Analysis (GSVA) method from the “GSVA” R package was applied to investigate the significantly altered pathways between high- and low-risk groups in the TCGA dataset and GSE14520. Furthermore, Gene set enrichment analysis (GSEA) was applied to evaluate the value of the signature by using diagnosis, metastasis risk and prognosis related gene sets (<http://software.broadinstitute.org/gsea/msigdb/index.jsp>).

Results

Construction of co-expression network and identification of metastasis-related module

Among the 378 screened datasets, GSE14520 was identified as the most idea dataset for constructing co-expression networks. It was performed on Affymetrix Human Genome U133A 2.0 Array and Affymetrix HT human Genome U133A Array, including 247 HCC samples with comprehensive clinical information

[24, 25]. The TCGA dataset was used to establish LASSO model, as it contains 365 HCC samples with complete gene expression profile and patient outcome information.

To satisfy normal distribution, the data cleaning process excluded 26 samples from the GSE14520 dataset for subsequent analysis (Additional file 1: Figure S1). To ensure a score-free network, the power of $\beta = 5$ (score-free $R^2 = 0.9$) was selected (Additional file 2: Figure S2). Based on 12 categories of clinical information, including metastasis risk (Fig. 1b), 14 modules were identified by using the “WGCNA” package in R (Fig. 1c). The blue module shows the highest correlation with the metastasis risk, and thus was subjected to further analysis (Fig. 1d, e). 16 hub genes that were closely connected to the blue module were identified for the following analysis (Fig. 1f).

Establishment of a 5-mRNA signature to predict overall survival

By using LASSO regression, a 5-parameter model was established after ten-time cross-validation to predict overall survival, including 4-aminobutyrate aminotransferase (ABAT), glycogen synthase 2 (GYS2), Hydroxyacid oxidase 1 (HAO1), hydroxyacylglutathione hydrolase (HAGH) and SEC14-like protein 2 (SEC14L2). The coefficients were -0.00504 , -0.0063 , -0.00165 , -0.00426 and -0.00147 , respectively (Additional file 3: Figure S3). They were all down-regulated in HCC tumor samples (Additional file 4: Table S1; Additional file 5: Figure S4; Additional file 6: Figure S5; Additional file 7: Figure S6; Additional file 8: Figure S7; Additional file 9: Figure S8). Using this 5-mRNA formula, the risk score for each patient was calculated (Fig. 2a). Then, the patients were divided into two groups based on the median of risk score (Fig. 2b). We found that patients with low risk score had a significantly higher survival rate ($p = 0.00013$) (Fig. 2c). Time-dependent ROC analysis has indicated that the prognostic accuracy of 5-mRNA signatures is 0.650 at 1 year, 0.627 at 3 years and 0.601 at 5 years (Fig. 2d).

Validation of the 5-mRNA signature for predicting overall survival, recurrence free survival and disease free survival

Overall survival and recurrence free survival data from GSE14520 and disease free survival data from TCGA LIHC were used to perform the validation. Based on GSE14520, patients with low risk score had a higher survival rate ($p = 0.0062$) (Fig. 2e, f, g). The prognostic accuracy of the 5-mRNA signatures is 0.650 at 1 year, 0.561 at 3 years, and 0.668 at 5 years (Fig. 2h). In TCGA LIHC, patients with low risk score had a elevated disease free survival rate (Fig. 2i, j, k) and the AUC was 0.689 at 1 year, 0.637 at 3 years and 0.597 at 5 years (Fig. 2l). Recurrence analysis was performed in GSE14520 and patients with high risk score had a higher recurrence rate ($p = 0.026$) (Fig. 2m, n, o). The AUC was 0.598 at 1 year, 0.546 at 3 years, and 0.609 at 5 years (Fig. 2p), respectively.

Prognostic value of the 5-mRNA signature in combination with essential clinical characteristics

To explore whether the prognostic value of the 5-mRNA signature is applicable to other clinical characteristics, stratification analysis was performed according to stage, grade, N stage, M stage, age and gender. We found that the 5-mRNA signature is still a clinically and statistically significant prognostic model (except in the female subgroup) (Fig. 3).

Clinical application of nomogram

To make the predictive model visual and quantized, we constructed a nomogram combining the 5-mRNA signature with other clinical characteristics, including age and stage, to predict the probability of 3- and 5-year overall survival (Fig. 4a). Calibration plots showed high performance of the nomogram compared with an ideal model (Fig. 4b, c). The decision curve analysis further confirmed the clinical usefulness of this nomogram (Fig. 4d, e).

Validation of pan-cancer overall survival and assessment of metastatic risk

To further broaden the clinical implications of our signature, pan-cancer validation of overall survival was analyzed. Consistent with the performance in HCC, the signature exhibited its prognostic value in breast cancer ($p = 0.028$), kidney renal clear cell carcinoma ($p = 0.021$), kidney renal papillary cell carcinoma ($p = 0.0011$), lung adenocarcinoma ($p = 0.0001$) and pancreatic ductal adenocarcinoma ($p = 0.017$) (Additional file 10: Figure S9), but not in cancers, such as bladder carcinoma, cervical squamous cell carcinoma and esophageal carcinoma (Additional file 11: Figure S10).

In respect to metastatic risk, the signature significantly distinguishes high/low metastasis survival rate. As for breast cancer, in GSE5327 cohort, patients with high risk score had a lower metastasis free survival rate ($p = 0.017$). Noticeably, lung metastasis survival rate had the same tendency ($p = 0.038$) (Fig. 5a, b). The results of metastasis free survival analysis were further confirmed in GSE20685, GSE24450 and GSE58644 datasets ($p = 0.043$, 0.00042 and 0.0091 , respectively) (Fig. 5c, d, e). The distant metastasis free survival of breast cancer in KM plotter database and metastasis free survival of GSE116918 (prostate cancer) were also significant ($p = 0.0014$ and 0.0038 , respectively) (Fig. 5f, g).

Alteration of biological pathways between high- and low-risk groups classified by the 5-mRNA signature

To identify the significant changes of biological pathways between high- and low- risk groups, we performed the gene set variation analysis (GSVA). Interestingly, metabolic changes, including bile acid metabolism, fatty acid metabolism, xenobiotic metabolism, heme metabolism, glycolysis and adipogenesis, were particularly prominent between these two groups, both in TCGA dataset and GSE14520 (Fig. 6a, b). GSEA analysis demonstrated that the high-risk group was enriched in gene sets which were positively related to metastasis/invasiveness and poor survival and low-risk group was enriched in gene sets negatively related to metastasis and poor survival (Fig. 6c).

Discussion

It is well-recognized that more than 90% of cancer mortality is caused by metastasis [26, 27]. High potential of metastasis is the main risk of death for postoperative HCC patients [28]. In the present study, we have established a signature of 5 metastasis-related mRNAs for prediction of tumour recurrence and patient survival in HCC.

These five genes, to our knowledge, are all related to metastasis in general but their role in HCC remains largely under investigated. ABAT encodes a key enzyme responsible for catabolism of principal inhibitory neurotransmitter gamma-aminobutyric acid (GABA). Loss of ABAT-mediated GABAergic system is associated with the aggressive behavior of basal like breast cancer (BLBC), providing potential prognostic indicators and therapeutic targets for BLBC [29]. GYS2 is a glycosyltransferase that catalyses the reaction of glucose to glycogen. It has been reported to be down-regulated in HCC and correlated with decreased glycogen content and unfavorable patient outcomes [30]. HAO1 catalyzes the oxidation of glycolate to glyoxylate within the peroxisomes of hepatocytes [31]. HAGH belongs to the main scavenging system of methylglyoxal, a potent prohe main scavenging system of methylglyoxalogen [32, 33]. It has been reported to involve in urological malignancies [34]. SEC14L2 is a cytosolic lipid-binding protein that has been reported to facilitate hepatitis C virus (HCV) replication [35, 36]. Collectively, these five metastasis-related genes play vital role in metabolism.

Liver is the central hub of metabolism, and metabolic reprogramming commonly occurs in cancer including in HCC [37, 38]. Metabolomic profiling has been widely explored for identification of potential biomarkers for early detection or prognosis in cancer [39, 40, 41, 42, 43]. In our study, interestingly, GSVA analysis revealed significant differences in metabolism between high- and low-risk groups, including bile acid metabolism, fatty acid metabolism and xenobiotics metabolism. Bile acids serve as signaling molecules in the liver that regulate lipid and glucose homeostasis [44]. They play an important role in the sustenance of inflammation, and thus the development and promotion of HCC [45, 46]. In cancer cells, 95% of fatty acids come from de novo synthesis [47]. In HCC, aberrant expression of lipogenic enzymes including fatty acid synthase is associated with tumor development and progression [48]. Our analysis revealing the dramatic metabolic changes between the different risk groups provides potential molecular targets for diagnosis, prognosis and therapeutic development.

In the field of biomarker development for HCC, it is well-recognized of the necessity to combine multiple molecules to enhance the sensitivity and specificity. For example, the most classical serological marker, AFP, was verified to achieve a higher sensitivity and specificity for diagnosing HCC when combined with AFP-L3 and des-gamma-carboxy prothrombin [49]. In this respect, our study has several advantages. Firstly, the 5-mRNA signature was established based on large dataset and innovative bioinformatics approaches. These are all metastasis-related genes identified by WGCNA, which indicates that the signature is functionally relevant to HCC progression and recurrence. Furthermore, the prognostic value of this signature was further validated in pan-cancer overall and metastatic survival. Secondary, with the combination of risk score and clinical factors, a nomogram was built to accurately predict the probability of overall survival in patients with HCC. Of note, there are still a few limitations of our study. Due to the

lack of data, we were not able to validate the prediction of metastasis in HCC. More importantly, prospective studies are required for a sound validation [50].

Conclusions

In summary, we have established a 5-mRNA signature and a nomogram for predicting the outcome and potential risk of metastasis and recurrence for HCC patients. This can be potentially developed as a promising tool for prognostic evaluation of HCC patients after surgical operation, but further validation is essentially required.

Abbreviations

HCC

hepatocellular carcinoma

WGCNA

weighted gene correlation network analysis

LASSO

least absolute shrinkage and selection operator

ROC

receiver operating characteristic

GSVA

gene set variation analysis

GSEA

gene set enrichment analysis

AFP

alpha-fetoprotein

GEO

Gene Expression Omnibus

TCGA

The Cancer Genome Atlas

AUC

area under curve

DCA

decision curve analysis

ABAT

4-aminobutyrate aminotransferase

GYS2

glycogen synthase

HAO1

hydroxyacid oxidase 1

HAGH
hydroxyacylglutathione hydrolase
SEC14L2
SEC14-like protein 2

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and supplementary files.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ZZ and JN made substantial contributions to conception and design of the research. ZZ carried out data collection and analysis. ZZ and JN wrote the paper. ZZ and JN edited the manuscript and provided critical comments. All authors read and approved the final manuscript

Acknowledgements

We thank the China Scholarship Council for funding PhD fellowships for Z. Zhou (No.201806270259).

References

1. Holm TM, Habashi JP, Doyle JJ, Bedja D, Chen Y, van Erp C, Lindsay ME, Kim D, Schoenhoff F, Cohn RD, et al. Noncanonical TGFbeta signaling contributes to aortic aneurysm progression in Marfan syndrome mice. *Science*. 2011;332(6027):358-61.
2. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet*. 2012;379(9822):1245-55.
3. Imamura H, Matsuyama Y, Tanaka E, Ohkubo T, Hasegawa K, Miyagawa S, Sugawara Y, Minagawa M, Takayama T, Kawasaki S, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *Journal of hepatology*. 2003;38(2):200-7.

4. Tang ZY. Hepatocellular carcinoma—cause, treatment and metastasis. *World journal of gastroenterology*. 2001;7(4):445-54.
5. Carr BI, Guerra V. Hepatocellular Carcinoma Extrahepatic Metastasis in Relation to Tumor Size and Alkaline Phosphatase Levels. *Oncology*. 2016;90(3):136-42.
6. Senthilnathan S, Memon K, Lewandowski RJ, Kulik L, Mulcahy MF, Riaz A, Miller FH, Yaghmai V, Nikolaidis P, Wang E, et al. Extrahepatic metastases occur in a minority of hepatocellular carcinoma patients treated with locoregional therapies: analyzing patterns of progression in 285 patients. *Hepatology*. 2012;55(5):1432-42.
7. Pinyol R, Montal R, Bassaganyas L, Sia D, Takayama T, Chau GY, Mazzaferro V, Roayaie S, Lee HC, Kokudo N, et al. Molecular predictors of prevention of recurrence in HCC with sorafenib as adjuvant treatment and prognostic factors in the phase 3 STORM trial. *Gut*. 2018;
8. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391(10127):1301-14.
9. Long J, Zhang L, Wan X, Lin J, Bai Y, Xu W, Xiong J, Zhao H. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *Journal of cellular and molecular medicine*. 2018;22(12):5928-38.
10. Yong KJ, Gao C, Lim JS, Yan B, Yang H, Dimitrov T, Kawasaki A, Ong CW, Wong KF, Lee S, et al. Oncofetal gene SALL4 in aggressive hepatocellular carcinoma. *The New England journal of medicine*. 2013;368(24):2266-76.
11. Han K, Zhou Y, Gan ZH, Qi WX, Zhang JJ, Fen T, Meng W, Jiang L, Shen Z, Min DL. p21-activated kinase 7 is an oncogene in human osteosarcoma. *Cell biology international*. 2014;38(12):1394-402.
12. Chaiteerakij R, Addissie BD, Roberts LR. Update on biomarkers of hepatocellular carcinoma. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*. 2015;13(2):237-45.
13. Zhang J, Chong CC, Chen GG, Lai PB. A Seven-microRNA Expression Signature Predicts Survival in Hepatocellular Carcinoma. *PloS one*. 2015;10(6):e0128628.
14. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 2008;9:559.
15. Sun M, Wu D, Zhou K, Li H, Gong X, Wei Q, Du M, Lei P, Zha J, Zhu H, et al. An eight-lncRNA signature predicts survival of breast cancer patients: a comprehensive study based on weighted gene co-expression network analysis and competing endogenous RNA network. *Breast Cancer Res Treat*. 2019;175(1):59-75.
16. Yuan L, Qian G, Chen L, Wu CL, Dan HC, Xiao Y, Wang X. Co-expression Network Analysis of Biomarkers for Adrenocortical Carcinoma. *Frontiers in genetics*. 2018;9:328.
17. Nakamura H, Fujii K, Gupta V, Hata H, Koizumu H, Hoshikawa M, Naruki S, Miyata Y, Takahashi I, Miyazawa T, et al. Identification of key modules and hub genes for small-cell lung carcinoma and large-cell neuroendocrine lung carcinoma by weighted gene co-expression network analysis of clinical tissue-proteomes. *PLoS One*. 2019;14(6):e0217105.

18. Zhou Z, Liu S, Zhang M, Zhou R, Liu J, Chang Y, Zhao Q. Overexpression of Topoisomerase 2-Alpha Confers a Poor Prognosis in Pancreatic Adenocarcinoma Identified by Co-Expression Analysis. *Dig Dis Sci*. 2017;62(10):2790-800.
19. R. T. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58(1):267-88.
20. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, Ye G, Deng H, Mou T, Cai S, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg*. 2018;267(3):504-13.
21. Zhang JX, Song W, Chen ZH, Wei JH, Liao YJ, Lei J, Hu M, Chen GZ, Liao B, Lu J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol*. 2013;14(13):1295-306.
22. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, Lin Y, Wang D, Yang X, Zheng Y, et al. DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. *Theranostics*. 2019;9(24):7251-67.
23. Backes Y, Elias SG, Groen JN, Schwartz MP, Wolfhagen FHJ, Geesing JMJ, Ter Borg F, van Bergeijk J, Spanier BWM, de Vos Tot Nederveen Cappel WH, et al. Histologic Factors Associated With Need for Surgery in Patients With Pedunculated T1 Colorectal Carcinomas. *Gastroenterology*. 2018;154(6):1647-59.
24. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer research*. 2010;70(24):10202-12.
25. Roessler S, Long EL, Budhu A, Chen Y, Zhao X, Ji J, Walker R, Jia HL, Ye QH, Qin LX, et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. *Gastroenterology*. 2012;142(4):957-66 e12.
26. Gupta GP, Massague J. Cancer metastasis: building a framework. *Cell*. 2006;127(4):679-95.
27. Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. *Nature medicine*. 2006;12(8):895-904.
28. Chang RM, Xiao S, Lei X, Yang H, Fang F, Yang LY. miRNA-487a Promotes Proliferation and Metastasis in Hepatocellular Carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2017;23(10):2593-604.
29. Besse A, Wu P, Bruni F, Donti T, Graham BH, Craigen WJ, McFarland R, Moretti P, Lalani S, Scott KL, et al. The GABA transaminase, ABAT, is essential for mitochondrial nucleoside metabolism. *Cell metabolism*. 2015;21(3):417-27.
30. Chen SL, Zhang CZ, Liu LL, Lu SX, Pan YH, Wang CH, He YF, Lin CS, Yang X, Xie D, et al. A GYS2/p53 Negative Feedback Loop Restricts Tumor Growth in HBV-Related Hepatocellular Carcinoma. *Cancer research*. 2019;79(3):534-45.
31. Williams E, Cregeen D, Rumsby G. Identification and expression of a cDNA for human glycolate oxidase. *Biochimica et biophysica acta*. 2000;1493(1-2):246-8.

32. Sousa Silva M, Gomes RA, Ferreira AE, Ponces Freire A, Cordeiro C. The glyoxalase pathway: the first hundred years... and beyond. *The Biochemical journal*. 2013;453(1):1-15.
33. Antognelli C, Del Buono C, Baldracchini F, Talesa V, Cottini E, Brancadoro C, Zucchi A, Mearini E. Alteration of glyoxalase genes expression in response to testosterone in LNCaP and PC3 human prostate cancer cells. *Cancer biology & therapy*. 2007;6(12):1880-8.
34. Antognelli C, Talesa VN. Glyoxalases in Urological Malignancies. *International journal of molecular sciences*. 2018;19(2)
35. Costa R, Todt D, Zapatero-Belinchon F, Schenk C, Anastasiou OE, Walker A, Hertel B, Timmer L, Bojkova D, Ruckert M, et al. SEC14L2, a lipid-binding protein, regulates HCV replication in culture with inter- and intra-genotype variations. *Journal of hepatology*. 2018;
36. Saeed M, Andreo U, Chung HY, Espiritu C, Branch AD, Silva JM, Rice CM. SEC14L2 enables pan-genotype HCV replication in cell culture. *Nature*. 2015;524(7566):471-5.
37. Macpherson AJ, Heikenwalder M, Ganai-Vonarburg SC. The Liver at the Nexus of Host-Microbial Interactions. *Cell host & microbe*. 2016;20(5):561-71.
38. Vander Heiden MG, DeBerardinis RJ. Understanding the Intersections between Metabolism and Cancer Biology. *Cell*. 2017;168(4):657-69.
39. Yang J, Xu G, Zheng Y, Kong H, Pang T, Lv S, Yang Q. Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences*. 2004;813(1-2):59-65.
40. Yang Y, Li C, Nie X, Feng X, Chen W, Yue Y, Tang H, Deng F. Metabonomic studies of human hepatocellular carcinoma using high-resolution magic-angle spinning ¹H NMR spectroscopy in conjunction with multivariate data analysis. *Journal of proteome research*. 2007;6(7):2605-14.
41. Patterson AD, Maurhofer O, Beyoglu D, Lanz C, Krausz KW, Pabst T, Gonzalez FJ, Dufour JF, Idle JR. Aberrant lipid metabolism in hepatocellular carcinoma revealed by plasma metabolomics and lipid profiling. *Cancer research*. 2011;71(21):6590-600.
42. Nahon P, Amathieu R, Triba MN, Bouchemal N, Nault JC, Ziol M, Seror O, Dhonneur G, Trinchet JC, Beaugrand M, et al. Identification of serum proton NMR metabolomic fingerprints associated with hepatocellular carcinoma in patients with alcoholic cirrhosis. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012;18(24):6714-22.
43. Beyoglu D, Imbeaud S, Maurhofer O, Bioulac-Sage P, Zucman-Rossi J, Dufour JF, Idle JR. Tissue metabolomics of hepatocellular carcinoma: tumor energy metabolism and the role of transcriptomic classification. *Hepatology*. 2013;58(1):229-38.
44. Torres DM, Williams CD, Harrison SA. Features, diagnosis, and treatment of nonalcoholic fatty liver disease. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*. 2012;10(8):837-58.
45. Jia W, Xie G, Jia W. Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. *Nature reviews Gastroenterology & hepatology*. 2018;15(2):111-28.

46. Sun L, Beggs K, Borude P, Edwards G, Bhushan B, Walesky C, Roy N, Manley MW, Jr., Gunewardena S, O'Neil M, et al. Bile acids promote diethylnitrosamine-induced hepatocellular carcinoma via increased inflammatory signaling. *American journal of physiology Gastrointestinal and liver physiology*. 2016;311(1):G91-G104.
47. Ookhtens M, Kannan R, Lyon I, Baker N. Liver and adipose tissue contributions to newly formed fatty acids in an ascites tumor. *The American journal of physiology*. 1984;247(1 Pt 2):R146-53.
48. Li L, Pilo GM, Li X, Cigliano A, Latte G, Che L, Joseph C, Mela M, Wang C, Jiang L, et al. Inactivation of fatty acid synthase impairs hepatocarcinogenesis driven by AKT in mice and humans. *Journal of hepatology*. 2016;64(2):333-41.
49. Johnson PJ, Pirrie SJ, Cox TF, Berhane S, Teng M, Palmer D, Morse J, Hull D, Patman G, Kagebayashi C, et al. The detection of hepatocellular carcinoma using a prospectively developed and validated model based on serological biomarkers. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2014;23(1):144-53.
50. Pesch B, Bruning T, Johnen G, Casjens S, Bonberg N, Taeger D, Muller A, Weber DG, Behrens T. Biomarker research with prospective study designs for the early detection of cancer. *Biochimica et biophysica acta*. 2014;1844(5):874-83.

Figures

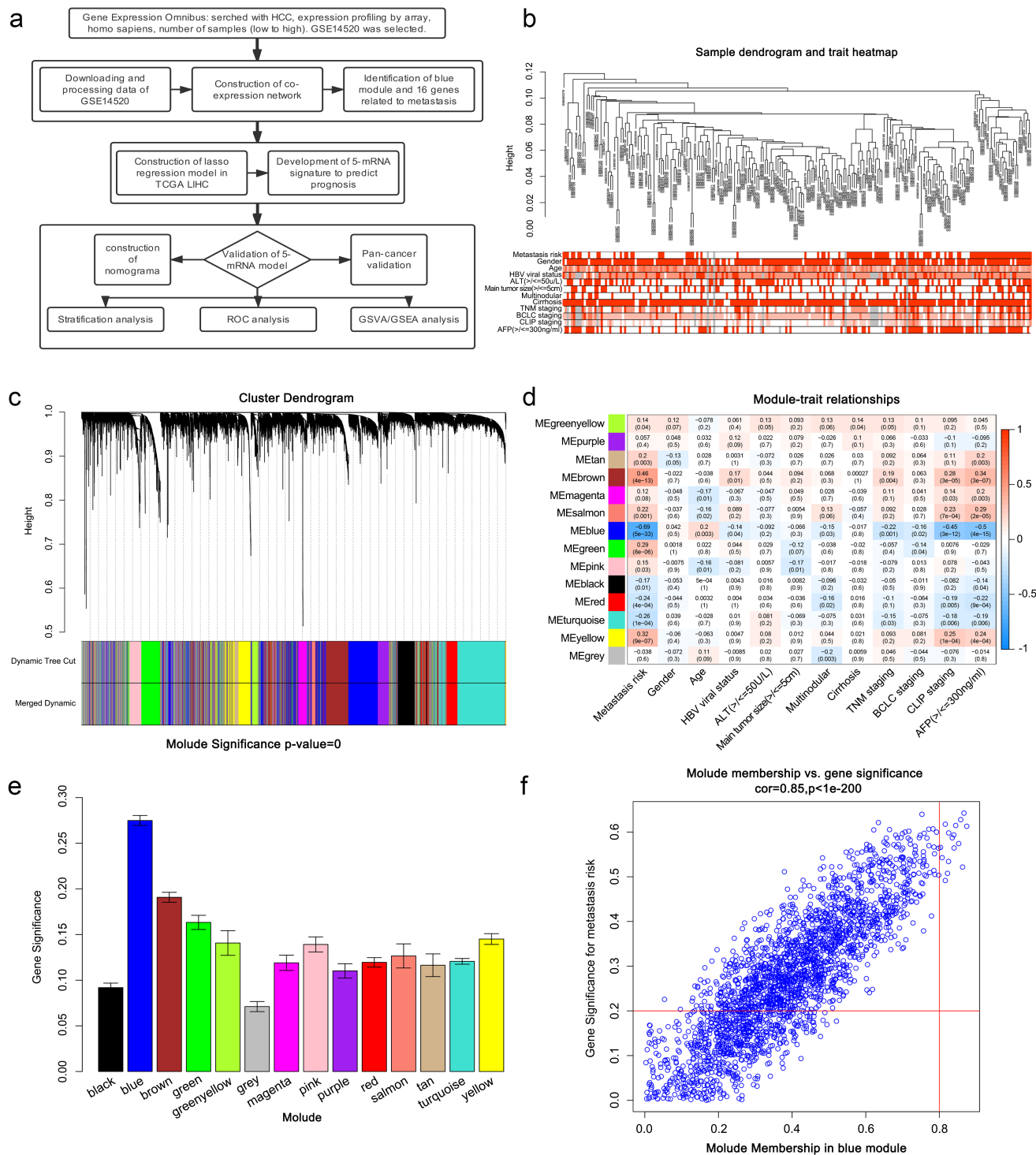


Figure 1

Construction of co-expression network and identification of hub module. (a) Flow diagram of analysis (b) Samples dendrogram and trait indicator (GSE14520). The clustering was based on the expression data of differentially expressed genes (DEGs) between tumor samples and non-tumor samples in HCC. (c) Dendrogram of all DEGs clustered based on a dissimilarity measure (1-TOM) (d) Heatmap of the

correlation between modules and traits. (e) Distribution of average gene significance and errors in the modules. (f) Scatter plot of module eigengenes in blue module.

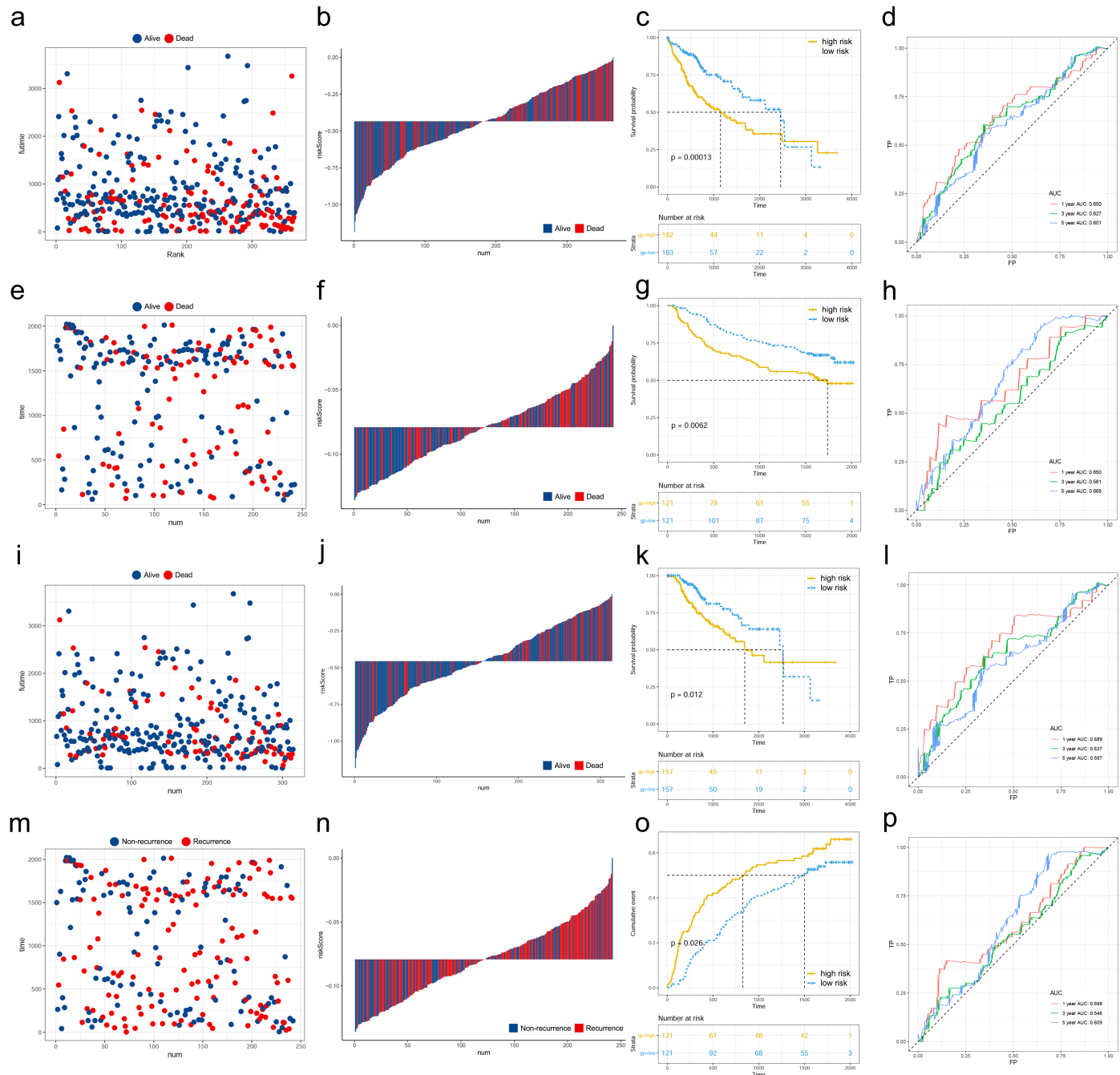


Figure 2

Risk score by the 5-mRNA signature, overall survival analysis, disease free survival analysis, recurrence free survival analysis and time-dependent ROC curve. (a/e) Distribution of patients with overall survival status in TCGA/GSE14520. (b/f) Two groups based on the risk score median in TCGA/GSE14520 (c/g) The overall survival analysis (d/h) Time-dependent ROC curve. (i/m) Distribution of patients with disease free survival status in TCGA/recurrence free survival status in GSE14520. (j/n) Two groups based on the

risk score median in TCGA. (k/o) The disease free survival analysis/recurrence free survival analysis. (l/p) Time-dependent ROC curve.

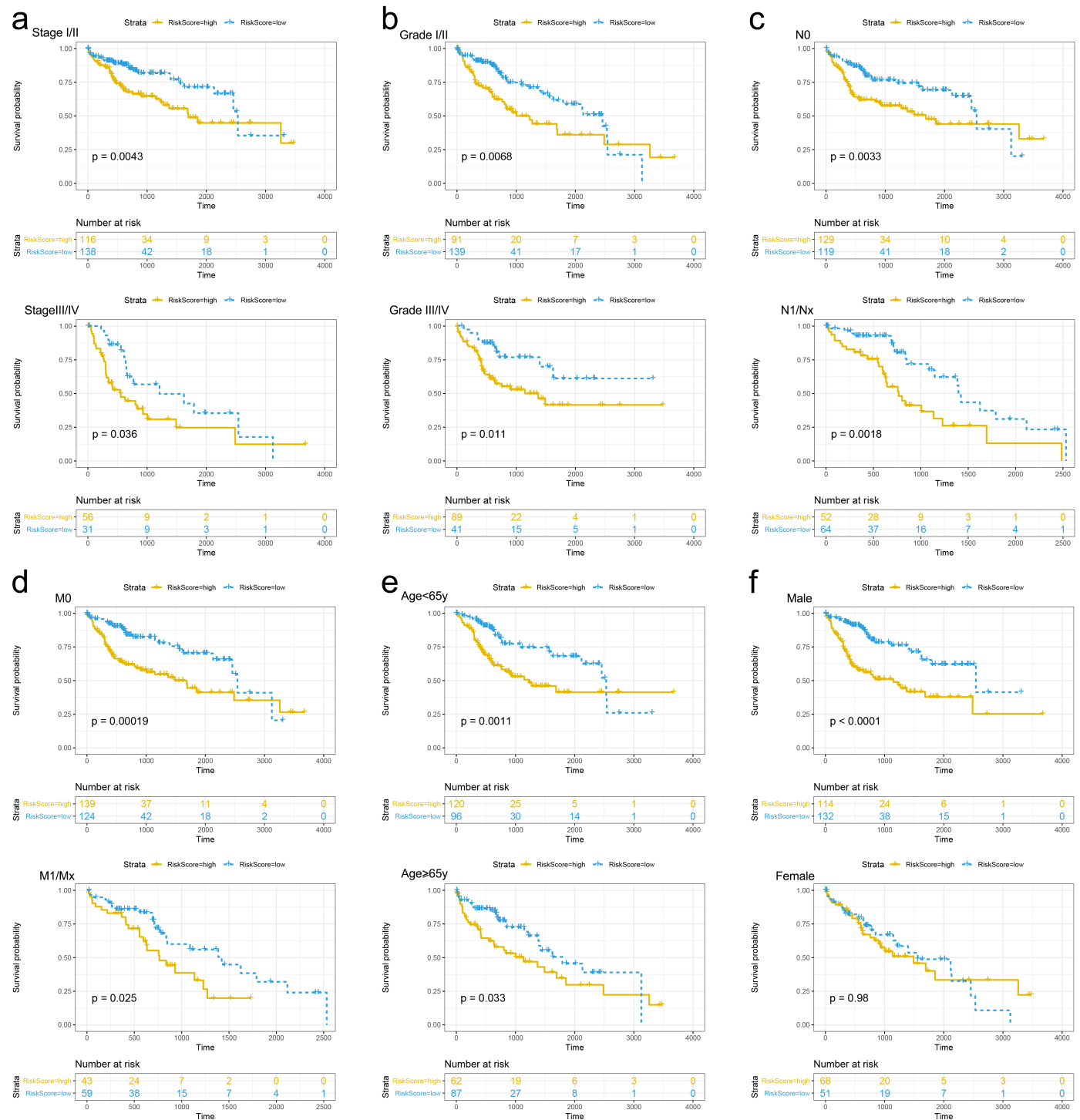
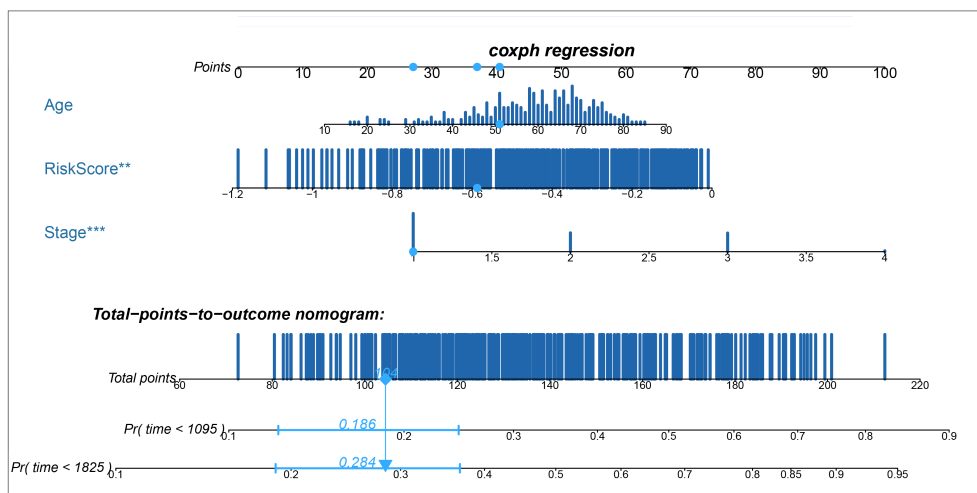


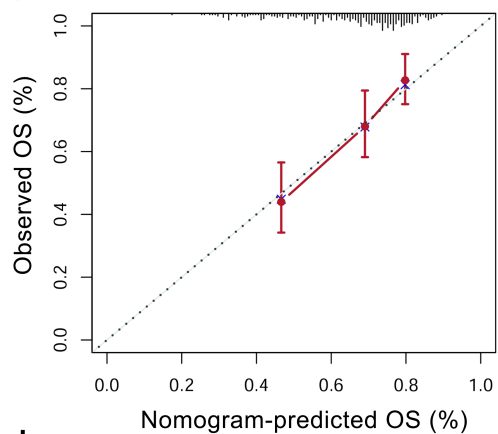
Figure 3

Prognostic efficiency of the 5-mRNA signature combined with the clinical features by stratified analysis. (a) The overall survival (OS) analysis for Stage I/II and Stage III/IV. (b) The OS analysis for Grade I/II and Grade III/IV. (c) The OS analysis for N0 and N1/Nx. (d) The OS analysis for M0 and M1/Mx. (e) The OS analysis for age < 65y and age ≥ 65y. (f) The OS analysis for male and female.

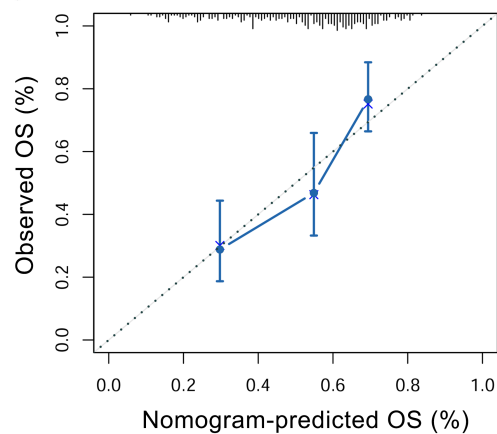
a



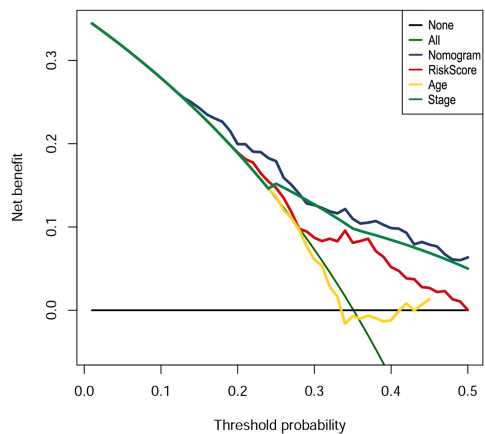
b



c



d



e

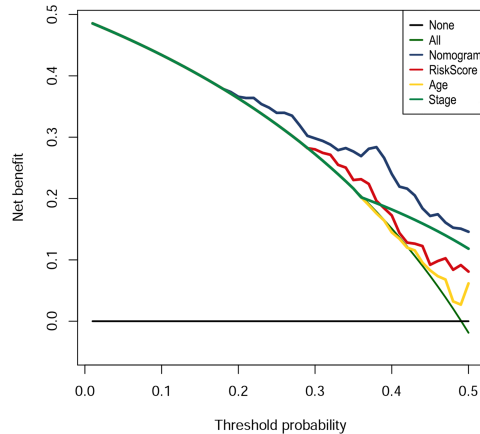


Figure 4

The nomogram to predict 3- and 5-year overall survival (OS) for HCC patients. (a) The nomogram for prediction proportion of patients with 3- or 5-year OS. Pr: prognostic risk. (b, c) The calibration plot for predicting patient 3- or 5-year OS. The X-axis represents the nomogram-predicted probability and the Y-axis represents the actual survival. (d, e) Decision curve analysis for assessment of the clinical utility of

the nomogram. The X-axis represents the percentage of the threshold probability and the Y-axis represents the net benefit.

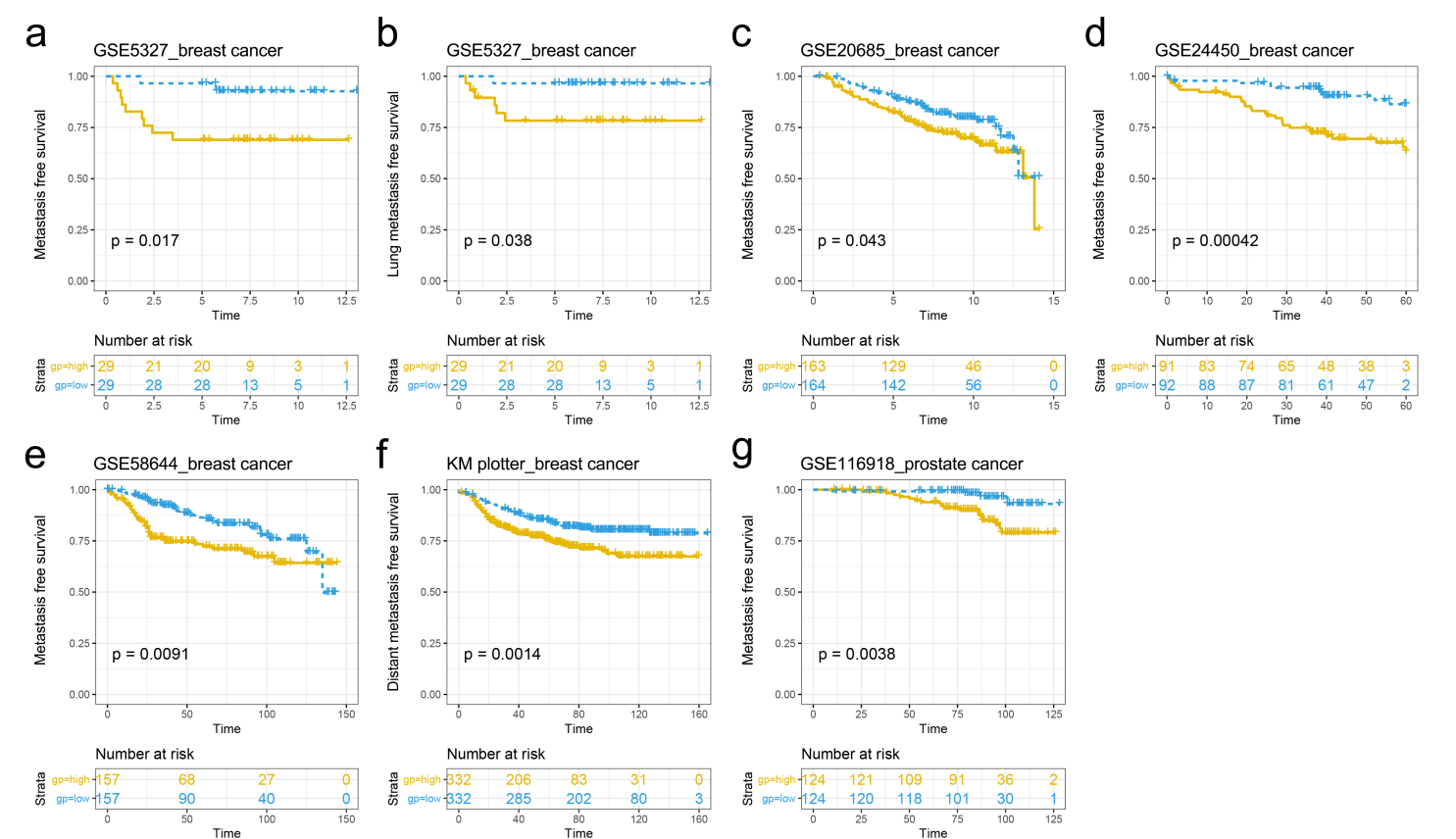


Figure 5

Prediction of metastasis free survival by the 5-mRNA signature in breast cancer and prostate cancer. (a) The metastasis free survival of breast cancer in GSE5327. (b) The lung metastasis free survival of breast cancer in GSE5327. (c) The metastasis free survival of breast cancer in GSE20685. (d) The metastasis free survival of breast cancer in GSE24450. (e) The metastasis free survival of breast cancer in GSE58644. (f) The distant metastasis free survival of breast cancer in KM plotter. (g) The metastasis free survival of prostate cancer in GSE116918.

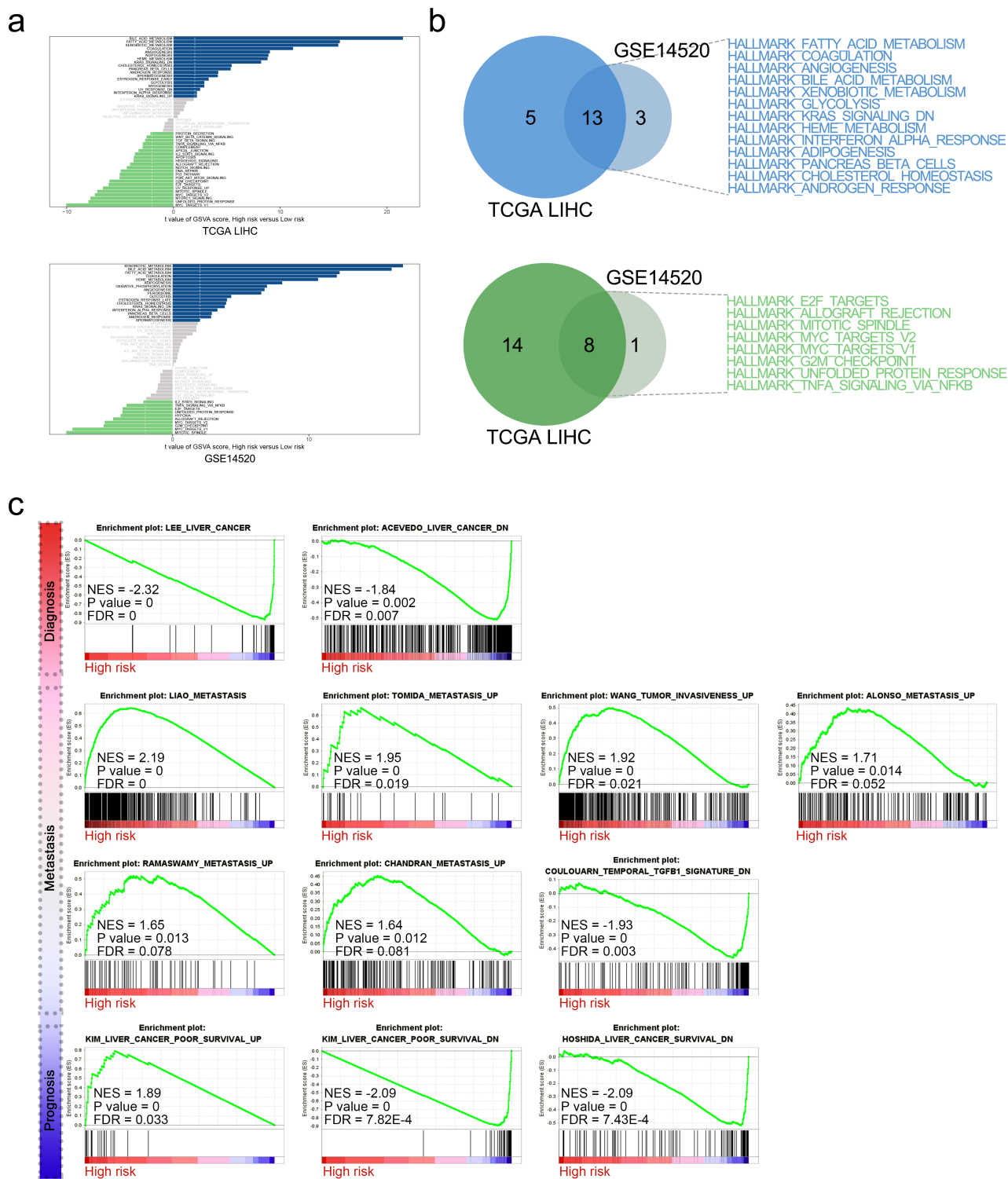


Figure 6

Gene set variation and enrichment analysis (GEVA and GSEA) (a) GSEA by using TCGA LIHC and GSE14520. (b) The common alteration in both datasets (c) GSEA by using TCGA.

Supplementary Files

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

- [Additionalfile5FigureS4.tif](#)
- [Additionalfile10FigureS9.tif](#)
- [Additionalfile11FigureS10.tif](#)
- [Additionalfile9FigureS8.tif](#)
- [Additionalfile7FigureS6.tif](#)
- [Additionalfile8FigureS7.tif](#)
- [Additionalfile1FigureS1.tif](#)
- [Additionalfile3FigureS3.tif](#)
- [Additionalfile2FigureS2.tif](#)
- [Additionalfile4TableS1.docx](#)
- [Additionalfiles.docx](#)
- [Additionalfile6FigureS5.tif](#)