Am J Transl Res 2021;13(5):4603-4613
www.ajtr.org /ISSN:1943-8141/AJTR0129067

Original Article

Prognostic gene biomarker identification in liver cancer by data mining

Gang Liu[1], Haitao Tang[1], Chen Li[1], Haiyan Zhen[2], Zhigang Zhang[2], Yongzhong Sha[3]

[1]School of Information Science and Engineering, Lanzhou University, Lanzhou, Gansu, China; [2]The First Hospital of Lanzhou University, Lanzhou, Gansu, China; [3]School of Management, Lanzhou University, Lanzhou, Gansu, China

Received December 27, 2020; Accepted February 19, 2021; Epub May 15, 2021; Published May 30, 2021

Abstract: Background: Liver cancer is a common cancer that enormously threatens the health of people worldwide. With the continuous advances of high-throughput gene sequencing technology and computer data mining technology, researchers can understand liver cancer based on the current accumulation of gene expression data and clinical information. Methods: We downloaded the TCGA data of liver cancer on the cancer-related website (https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/), comprising 438 patients and 20,530 genes. After removing some patients with missing survival data, we collected 397 patients' samples. Our data were collected from a public database without real patient participation. While matching the patient samples in the gene expression spectrum, we attained 330 samples with primary tumors and 50 samples with normal solid tissue. Results: After the 330 tumor tissue samples were randomized into two equal-numbered groups (one is a training set, and the other is a test set), we selected 26 gene biomarkers from the training set and validated them in the test set. Based on the selected 26 gene biomarkers, RBM14, ALG11, MAG, SETD3, HOXD10 and other 26 genes were considered independent risk factors for the prognosis of liver cancer, and genes such as GHR significantly affect human growth hormone for liver cancer. The findings discovered that low-risk patients survived remarkably better than the high-risk patients ($P<0.001$), and the area under the curve (AUC) of receiver operating characteristic curve (ROC) was greater than 0.5. Conclusion: Our numerical results showed that these 26 gene biomarkers can be used to guide the effective prognostic therapy of patients with liver cancer.

Keywords: LASSO, liver cancer, gene biomarkers

Introduction

Liver cancer is a common malignant tumor, which is classified into primary and secondary. The liver may be impacted by primary liver cancer, which occurs in the liver, or by cancer that originates from other parts of the body and then spreads to the liver. But most liver cancer is secondary or metastatic, indicating that it originates from elsewhere in the body. Primary liver cancer occurs in the liver. About 2% of cancers in the U.S, but up to half of all cancers in some undeveloped countries, result from the epidemic of hepatitis caused by infectious viruses that is susceptible to liver cancer. In the U.S., men with primary liver cancer is twice in number as of women [1]. In China, 2010 saw 358,840 new patients diagnosed with liver cancer and 312,432 died of this disease. Based on China's population at that time, incidence and mortality rate were 0.0027% and 0.023%, especially in underdeveloped rural and western regions. China's liver cancer incidence and mortality rates are much higher than the world average [2].

Currently, the etiology and molecular mechanism of primary disease are not yet clear, but it is known that there are many influencing factors. Hepatitis B virus infection, aflatoxin intake, and long-term consumption of unclean water are the main causes of liver cancer in the world [3]. With the change of diet habits and neonatal hepatitis B vaccinations, the exposure of risk factors is decreasing, but liver cancer incidences are still increasing [4].

Medical research and drug development for liver cancer are of great significance to people all over the world. Presently, targeted drugs [5] have wide application prospects in liver cancer treatment compared with traditional treatment methods, and they have incomparable advantages by targeting abnormally expressed genes rather than targeting normally expressed genes [6, 7]. However, it is important and difficult to select the safest and most appropriate genes for targeted gene therapy. In this paper, we hope to provide some technical support for selecting the appropriate genome by data mining and analysis of genetic biomarkers of hepatocellular carcinoma.

The advances of gene chip technology and next-generation sequencing technology [8] assist in mastering liver cancer research from the genomic perspective and proposing effective gene biomarkers, which provides effective scientific evidence for drug development and clinical decision-making. In the meantime, the individualized medical treatment and precision medical care have been constantly modified [9]. Through high throughput gene expression experiments, we generally find that there are more observed values than samples, and traditional statistical methods have higher dimensional problems. This study used least absolute shrinkage and selection operator (LASSO) to perform data mining on genomic data to minimize the instability caused by high dimensional data [10].

At first, the Cancer Genome Atlas (TCGA) data of liver cancer were obtained from cancer samples, and 15 high-risk genes and 11 low-risk genes were calculated by Cox regression analysis. Based on the linear combination of gene expression and Cox regression coefficient, we allocated the patients into a high-risk group and a low-risk group [11-14] by the value of each patient's prognosis index. Using survival analysis, we further verified whether the prognostic genes analyzed by mining can significantly distinguish the patients of the two types.

Materials and methods

Data collection

We downloaded the TCGA data of liver cancer on the cancer-related website (https://genome-cancer.ucsc.edu/proj/site/hgHeatmap/), comprising 438 patients and 20,530 genes. After removing some patients with missing survival data, we collected 397 patients' samples [15-17]. All data were collected from a public database without real patient participation.

Data preprocessing

By eliminating 5 incomplete gene data from 397 patients' samples, the remaining 392 samples were obtained and further divided into tumor tissue (336) and normal tissue (56) according to the characteristics of clinical data. After matching the patient samples in the gene expression spectrum, we attained 330 samples with primary tumors and 50 samples with normal solid tissue.

Finally, we equally randomized 330 patient samples with primary tumors into two groups (one for a training set, and the other for a test set) using the "sample" function.

Clinical data analysis

Many factors in clinical variables usually have an impact on the prognosis of patients with liver cancer, so many clinical variables should be considered as influencing factors. Based on previous research, some risky clinical factors including age, sex, cancer status, family cancer history, and pathologic staging were used for analyzing the clinical factors relevant to the subsequent data mining analysis [18-21].

Analysis of gene data

First of all, we used the "Limma" package of R language (version number 3.4.0) to analyze the gene differential expression of 165 training set samples and 50 normal tissue samples. In accordance with the adjusted *P* value (adj.P.val <0.001) and fold change (logFC ≥ 1 or logFC ≤ -1), we screened 5,075 differentially expressed genes [22, 23].

We then used the "survival" package in R to carry out Cox regression analysis of 5,075 genes and calculated the Cox coefficient, the hazard ratio (HR) value, and the *P* value of each gene using the Wald test. We screened 475 genes ($P<0.05$) that had significant correlation with the survival of the patients [24-26].

Finally, the 475 genes were treated by the LASSO method again in the R "glmnet" soft-

Table 1. Clinical information statistics

Factors	Patients (Death)	Log-rank test *P* value	HR (95% CI)
Age		0.109	1.309
> 63	89/198		[0.9406, 1.821]
≤ 63	860/199		
Gender		0.5742	0.9716
Male	88/262		[0.6937, 1.361]
Female	261/135		
Cancer_status			
Unknown	13/24		0.5608
With tumor	82/152	0.0009918	[0.3956, 0.795]
Tumor-free	54/221		
Family_cancer_history			
Unknown	13/49		0.6195
Yes	69/137		[0.4415, 0.8693]
No	67/211	0.005174	
Pathologic_T			
Unknown	0/1		
T0	0/1		
TX	1/1		1.9636
T1	61/200	7.052e-05	[1.399, 2.756]
T2	35/101		
T3	40/78		
T4	12/15		

ware package to gain more critical genes. After 90% off cross-validation, 26 risk genes, which closely correlate with survival, were finally obtained [27].

Prognostic index calculation

The prognostic index is an important indicator of the integration of risk genes, and provides a prognostic index (PI) value for each patient with liver cancer. PI is calculated by linearly fitting the product of the expression and the coefficient corrected by LASSO of each gene. The formula of the prognostic index is as follows:

$$P_I=\beta_1X_1+\beta_2X_2+...+\beta_iX_i+...+\beta_aX_n, \quad (1)$$

where X_i indicates the ith gene expression, and β_i is the coefficient corrected by LASSO of the ith gene.

Dataset verification

After the whole process was carried out on the training set sample, 26 gene biomarkers were finally screened out. Then, 165 samples of the test set were used to verify the feasibility of the 26 gene biomarkers.

Statistical methods

We analyzed the survival of patients with the use of Kaplan-Meier curve, and performed significance analysis using the double tail log-rank test. The *P* value was set to less than 0.05 as a threshold to judge whether the difference was significant [28].

We did Univariate survival analysis and multivariate Cox regression analysis by means of the survival package of R. The ROC curve was pictured with the use of the survival ROC package of R based on its glmnet software package. LASSO analysis was carried out by using 90% off cross-validation and 1000 iteration operation.

Experimental result

Statistical analysis of clinical variables

From the clinical materials of 330 patient samples, we took the patients' age, gender, cancer status, family history of cancer, and pathologic staging as the single variable. We carried out the Cox regression analysis by means of the "survival" package in R and calculated the *P* value of the Log-rank test and the HR value of each clinical factor (**Table 1**). Based on the selected 26 gene biomarkers, 26 genes such as RBM14, ALG11, MAG, SETD3 and HOXD10 are individual risk factors in the prognosis of liver cancer, and genes such as GHR significantly affect the human growth hormone of liver cancer. We developed the survival curves of each group, as shown in **Figures 1**-**3** and summarized the collected data in **Table 2**.

Analysis of gene prognostic index

By linear fitting the product of expression and regression coefficient of the 26 genes in each sample, we gained the prognostic index (PI) for each patient and ranked the patient indices from small to large. In accordance with the median PI value, the patients were allocated into high-risk patients and low-risk patients (see **Figure 4A**). We pictured the survival curve of both groups by Kaplan Meier (**Figure 4B**). When we took 5 years as the survival time, we

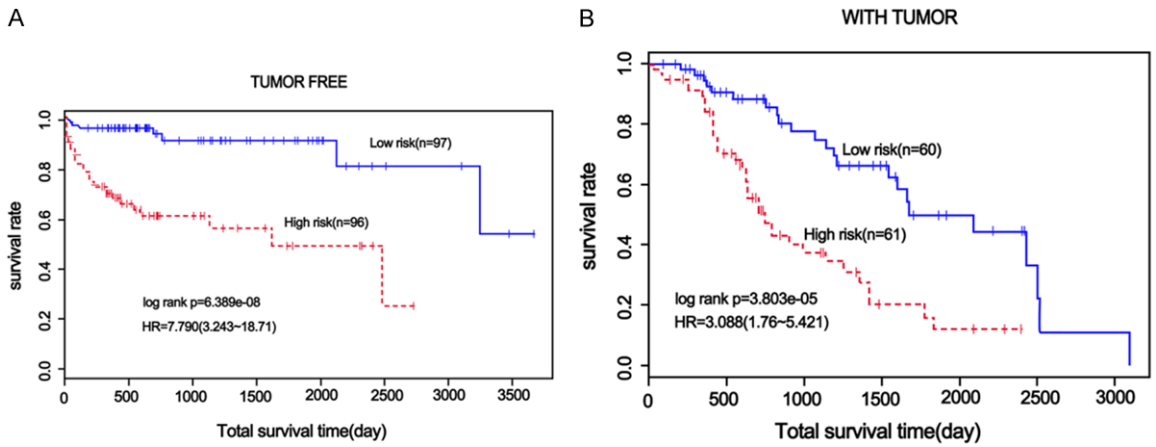

Figure 1. Survival curves of tumor and tumor free groups.

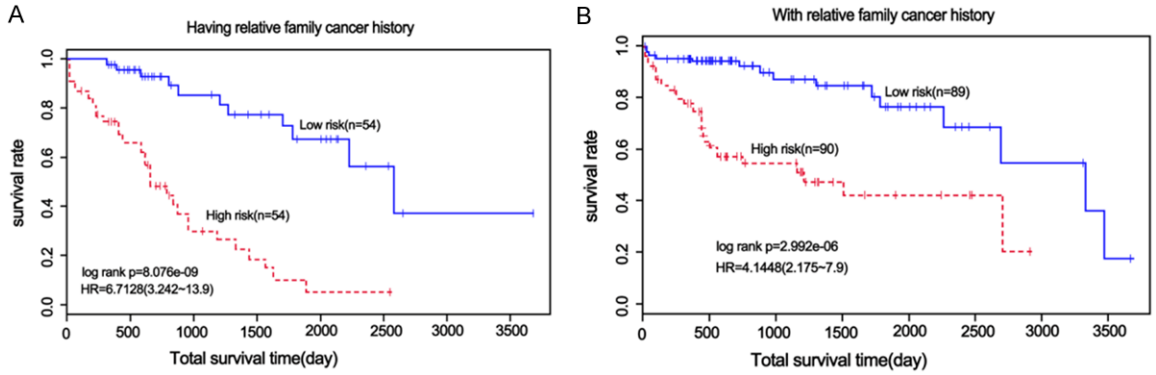

Figure 2. Survival curves of relative family cancer history and without relative family cancer history groups.

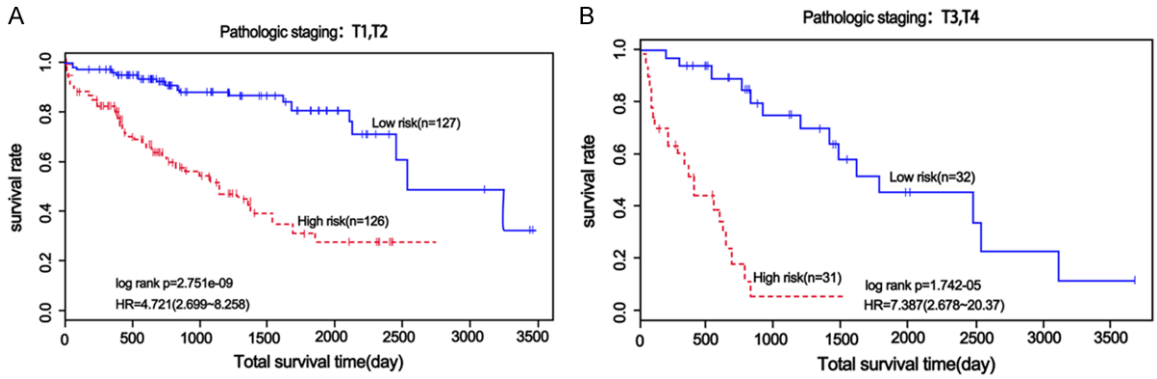

Figure 3. Survival curves of different pathologic stagings.

first drew the ROC curve using the "survival ROC" package (see **Figure 4C**). **Figure 4D** shows the result of the heatmap of 26 genes in high- and low-risk patients. We used the 26 genetic biomarkers to validate the test set samples, so as to further verify their accuracy (results shown in **Figure 5**).

Functional analysis of prognostic genes

For these 26 gene prognostic biomarkers, functions or pathways of these biomarkers were analyzed using the online gene ontology analysis tool DAVID [29, 30], and the biological function was shown in **Figure 6**. We imported these

Table 2. Gene list

Gene name	Full name	(HR, 95% CI)	Coefficient of Cox	*P* value
Risk gene				
RBM14	RNA binding motif protein 14	1.00043	0.048972159	0.000424
ALG11	Alpha-1,2-mannosyltransferase	1.020704	0.042378391	0.002449
MAG	Myelin-associated glycoprotein	1.011657	0.026155322	0.005336
SETD3	SET domain containing 3	1.010442	0.025303277	0.04619
HOXD10	Homeobox D10	1.023025	0.023870329	0.000146
NAB1	NGFI-A binding protein 1	1.001622	0.021278692	0.010548
HOXD1	Homeobox D1	1.02395	0.020723858	0.000112
VPS33A	Vacuolar protein sorting-associated protein 33A	1.000004	0.02019773	0.016008
SPA17	Sperm autoantigenic protein 17	1.018591	0.01922508	0.007574
FAM18B2	Golgi apparatus membrane protein TVP23 homolog C	1.016758	0.017435934	0.042599
HSD17B12	Hydroxysteroid 17-beta dehydrogenase 12	1.012643	0.017412119	0.04198
AKIRIN1	Akirin 1	1.007642	0.016965499	0.017409
XPR1	Xenotropic and polytropic retrovirus receptor 1	1.001022	0.016692604	0.004736
ZDBF2	zinc finger DBF-type containing 2	1.000684	0.015541813	0.005524
SETP3	Septin 3	1.005063	0.013854683	0.003262
Protective gene				
CFHR3	Complement factor H related 3	0.998525	-0.007495812	0.000474
GHR	Growth hormone receptor	0.997015	-0.008937969	0.004232
PSD4	Pleckstrin and Sec7 domain containing 4	0.990101	-0.011994413	0.003506
AHSA2	Activator of heat shock protein ATPase 2	0.997407	-0.014005068	0.017464
PZP	Pregnancy zone protein	0.99091	-0.01422213	0.032387
SOCS2	Suppressor of cytokine signaling 2	0.99	-0.014911698	0.000493
MGMT	Methylated-DNA--protein-cysteine methyltransferase	0.976306	-0.0179435	1.96E-05
KLRK1	Killer cell lectin like receptor K1	0.9918	-0.018454796	0.01954
VPS39	Vam6/Vps39-like protein	0.988445	-0.021095758	0.032571
C9orf163	Chromosome 9 open reading frame 163	0.981049	-0.024429344	0.021895
LOC285359	Unknown	0.990937	-0.032193194	0.034797

26 genes into the online gene analysis tool STRING to analyze the related gene functions (**Figure 7**). These growth-related signaling pathways are strongly correlated with proliferation and metastasis of cancer cells (**Table 3**) [32]. In the analysis of cellular components, we found that different combinations of genes make up different cells (**Table 4**).

Discussion

In this study, diverse statistical analysis methods, consisting of LASSO regression, univariate survival analysis, multivariate Cox proportional risk regression model, and ROC curve analysis, were used for mining the gene expression profile of liver cancer. In particular, with the use of the supervised clustering analysis, we screened the patients with high risk or low risk on the training set and test set. The results showed that the method in this paper is highly accurate.

By linear fitting the product of expression and regression coefficient of the 26 genes in each sample, we gained the prognostic index (PI) for each patient and ranked the patient indices from small to large. In accordance with the median PI value, the patients were allocated into high-risk patients and low-risk patients, where a lower PI value indicated that patients were at lower risk of death. With the use of Kaplan Meier, we calculated that the *P* value of the Log-rank test was less than 0.001, and HR was equal to 11.079, indicating that the two groups of patients have significant differences in survival time. We judged the merits and demerits of the model constructed by the 26 gene biomarkers according to the AUC value. The findings discovered that AUC was 0.912 (AUC > 0.5 means a well-behaved model) indicating that the model constructed by these 26 gene biomarkers performs very well. Based on the experimental results of the training set sample data, we found that 26 gene prognostic

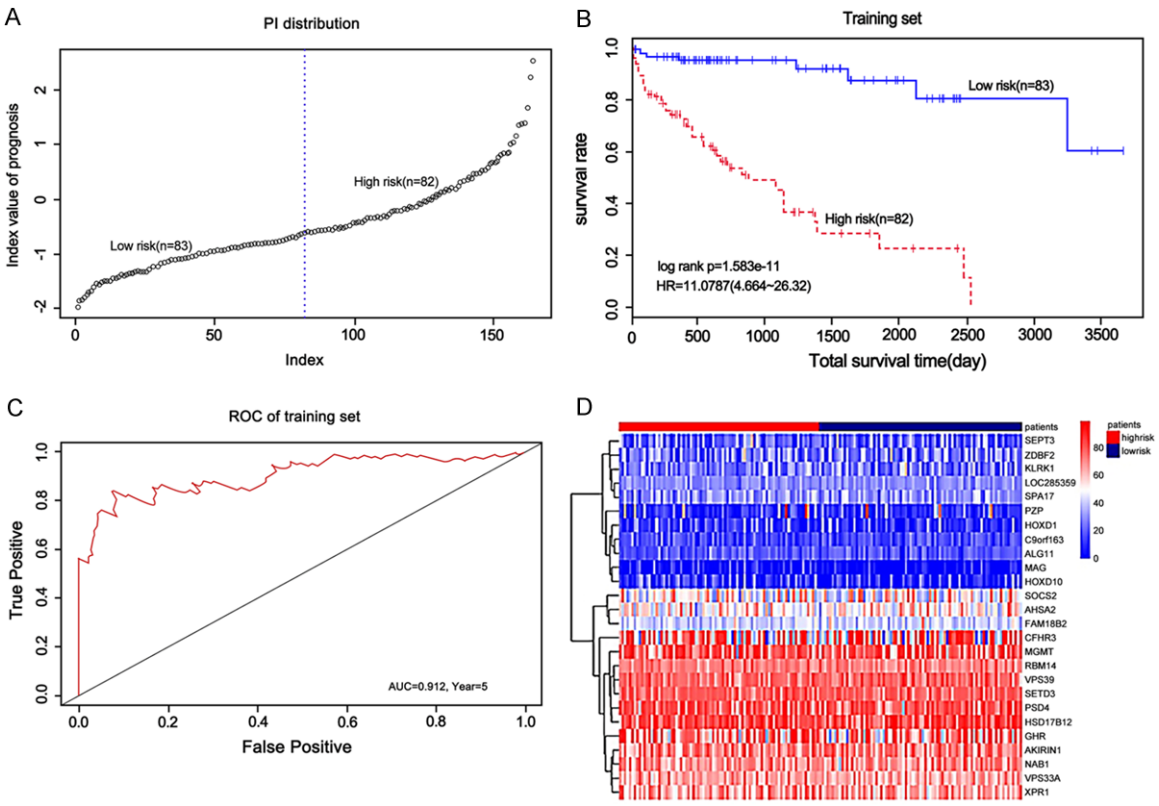

Figure 4. Survival curve, ROC and heatmap of high-risk and low-risk patients.

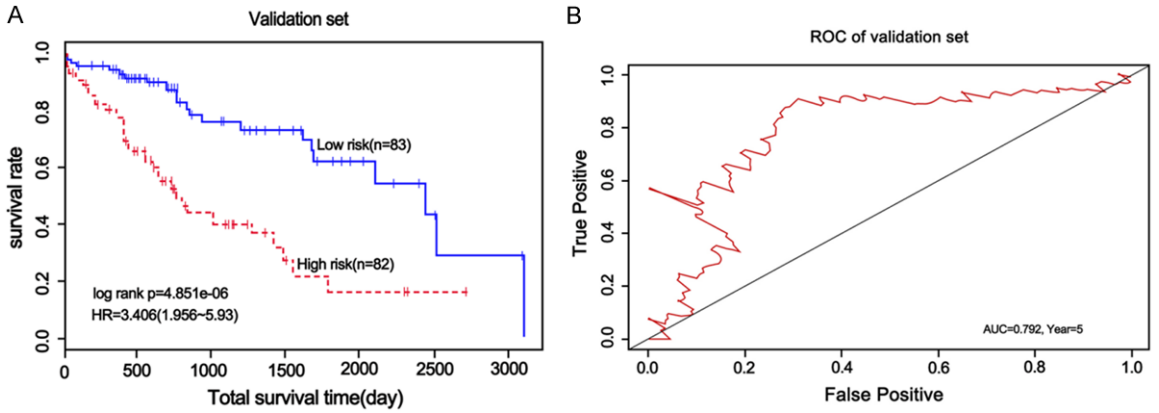

Figure 5. The accuracy of the test results.

biomarkers could be used to significantly classify HCC (hepatic cell carcinoma) patients into two parts: high risk and low risk. Based on these 26 biomarkers, 165 patients were assigned to two groups in accordance with Kaplan-Meier, where the *P* value of the Log-rank test was less than 0.001, and the risk value of HR was 3.406. A lower area AUC (0.792) in the ROC receiver's curve, indicating that the 26 gene biomarkers can still separate patients into high risk and low risk based on the test set sample.

For these 26 gene prognostic biomarkers, functions or pathways of these biomarkers were analyzed using the online gene ontology analysis tool DAVID [29, 30]. We found that these 26 gene biomarkers are mainly involved in skeletal

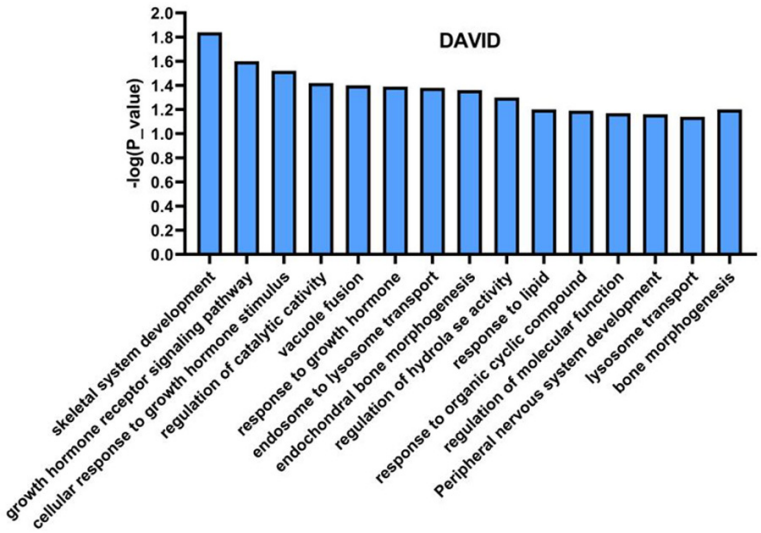

Figure 6. The analysis of functions or pathways of these biomarkers using the online gene ontology analysis tool DAVID29FFF.

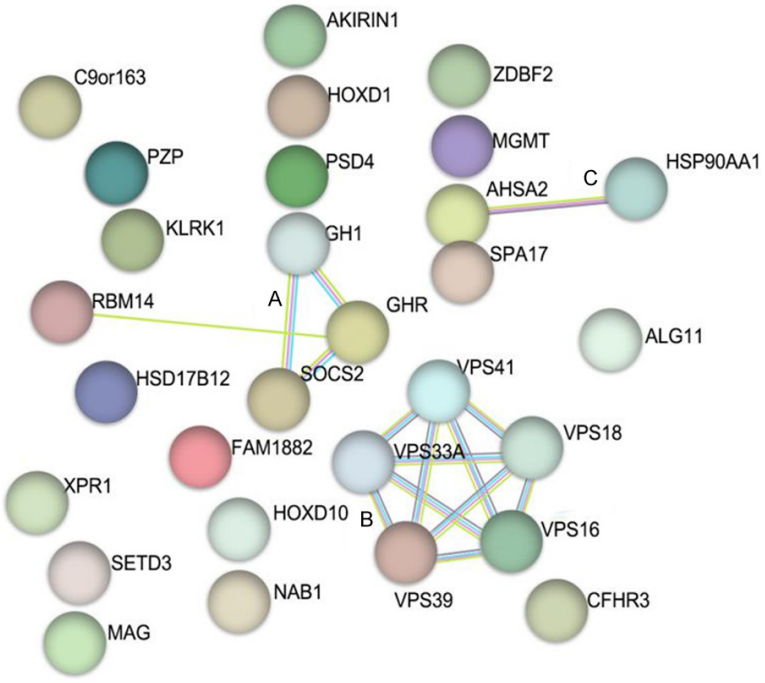

Figure 7. The analysis of the related gene functions using online gene analysis tool STRING.

system growth, growth hormone receptor signaling pathway, cell response to growth hormone stimulation, regulation of catalytic activity, vacuolar fusion, growth hormone reaction, and lysosomal transport in the inner body. To further understand the relationship between these biomarkers and the functions or pathways involved, we imported these 26 genes into the online gene analysis tool STRING to analyze the related gene functions [31]. The A pathway in **Figure 7** includes GH1, GHR, and SOCS2 genes, which are closely associated with human growth hormone that impacts liver cancer [31]. The B pathway in **Figure 7** includes vps-16, vps18, and vps41, which are directly related to the formation of SNARE protein. Therefore, vps16, vps18, and vps-41 may provide a new direction for the treatment of liver cancer [31]. The C pathway in **Figure 7** includes HSP90AA1 and AHSA2, which are mainly related to ATPase [32].

The development pathway results indicate that these genes are mainly enriched in the combination of soluble N-ethyl maleimide sensitive protein receptors. Other genes mainly participate in the pathway of growth hormone receptor and the cell response pathway stimulated by growth hormone. These growth-related signaling pathways are strongly correlated with proliferation and metastasis of cancer cells [32]. In the analysis of cellular components, we found that different combinations of genes make up different cells. These gene products usually form a number of protein complexes to adsorb on the cell membrane and then merge into the nuclear body.

According to the results of STRING, GH1, GHR, and SOCS2 were associated with the epidermal growth factor receptor, which was related to liver cancer development [33, 34]. Additionally, vps16, vps18, and vps41 were associated with formation of SNARE. SNARE protein is essential to the vesicle trafficking between cellular com-

Table 3. Biological process (GO)

Pathway ID	Pathway description	Observed gene count	False discovery rate	Matching proteins in your network (labels)
GO.0035542	Regulation of SNARE complex assembly	3	0.00197	VPS16, VPS18, VPS41
GO.0060396	Growth hormone receptor signaling pathway	3	0.0202	GH1, GHR, SOCS2
GO.0060397	JAK-STAT cascade involved in growth hormone signaling pathway	3	0.0202	GH1, GHR, SOCS2
GO.0071378	Cellular response to growth hormone stimulus	3	0.0202	GH1, GHR, SOCS2

Table 4. Cellular component (GO)

#Pathway ID	Pathway description	Observed gene count	False discovery rate	Matching proteins in your network (labels)
GO.0030897	HOPS complex	4	5.33E-09	VPS16, VPS18, VPS39, VPS41
GO.0005770	late endosome	5	0.00496	VPS16, VPS18, VPS33A, VPS39, VPS41
GO.0031902	Late endosome membrane	4	0.00496	VPS16, VPS18, VPS33A, VPS39
GO.0005765	Lysosomal membrane	5	0.00713	VPS16, VPS18, VPS33A, VPS39, VPS41
GO.0044440	Endosomal part	5	0.0435	VPS16, VPS18, VPS33A, VPS39, VPS41

partments and the plasma membrane, and the key roles of SNAREs are commonly to lend specificity for vesicle trafficking and mediate the membrane fusion process [35-38]. Because vesicle transport is involved in many important cell life events, vps16, vps18, and vps41 may provide new directions for liver cancer treatment.

According to string function analysis, we obtained a network of genes including three pathways. Apart from our above conclusion description, the results also illustrated that HSP90AA1 is a molecular chaperone that aids in protein folding and quality control for many "client" proteins. HSP90AA1 operates as a dimer and has intrinsic ATPase activity. In addition, AHSA2 is a co-chaperone that stimulates HSP90 ATPase activity, which significantly affects liver cancer [39-42].

Among the 26 genes screened in this study, some scholars have confirmed that seven genes, including MAGG, SETD3, HOXD10, HOXD1, GHR, SOCS2, and MGMT, are closely related to liver cancer. For example, if HOXD-AS1 expression rises, the overall survival of patients with liver cancer dramatically decreases, which is an independent prognostic indicator [43]. The miR-224/HOXD10/p-PAK4/MMP-9 signaling pathway makes contribution to metastasis regulation and cell invasion [44, 45]. HOXD10 restrains the proliferation of HCC cells by limiting ERK signal [46]. However, if SOCS2 expression in patients with HCC decreases, the proliferation and metastasis of HCC cells will accelerate, indicating that SOCS2 is crucial in inhibiting HCC metastasis [47-49]. EGFR-AS1 (target of GHR) can improve the ability of metastasis and proliferation of hepatoma cells and accelerate further liver cancer development. Moreover, it has been found that EGFR-AS1 can affect the growth cycle of hepatoma cells, so EGFR-AS1 (target of GHR) is a very important prognostic factor and provides a possible treatment for liver cancer [50]. It is possible that other gene biomarkers screened by this method have not been reported in the relevant literature of liver cancer disease, indicating that these potential targets have not yet been found. Further experimental exploration and clinical examination are necessary to understand the role of these gene biomarkers. More stable genetic markers are expected to be found by this method, which offers a more scientific reference in support of drug development and clinical decision-making.

Acknowledgements

This study was sponsored by the Key Project of China Ministry of Education for Philosophy and Social Science: Big Data Driven Risk Research on City's Public Safety (Grant No. 16JZD023) and the Fundamental Research Funds for the Central Universities: Big Data Driven Risk Prewarning Research on City's Public Safety (Grant No. 17LZUJBWZD012).

Disclosure of conflict of interest

None.

Address correspondence to: Gang Liu, School of Information Science and Engineering, Lanzhou University, Lanzhou, Gansu, China. Tel: +86-18109-495380; E-mail: Dr_Liu_Gang@126.com

References

[1] Jin J, Chen S, Wang D, Chen Y, Wang Y, Guo M, Zhou C and Dou J. Oroxylin A suppresses influenza A virus replication correlating with neuraminidase inhibition and induction of IFNs. Biomed Pharmacother 2018; 97: 385-394.

[2] Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, Bonaventure A, Valkov M, Johnson CJ, Estève J, Ogunbiyi OJ, Azevedo E Silva G, Chen WQ, Eser S, Engholm G, Stiller CA, Monnereau A, Woods RR, Visser O, Lim GH, Aitken J, Weir HK and Coleman MP; CONCORD Working Group. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 2018; 391: 1023-1075.

[3] Duan XY, Zhang L, Fan JG and Qiao L. NAFLD leads to liver cancer: do we have sufficient evidence? Cancer Lett 2014; 345: 230-4.

[4] Zuo T, Zheng R, Zeng H, Zhang S and Chen W. Analysis of liver cancer incidence and trend in China. Zhonghua Zhong Liu Za Zhi 2015; 37: 691-6.

[5] Sia D, Villanueva A, Friedman SL and Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. Gastroenterology 2017; 152: 745-761.

[6] Li L and Wang H. Heterogeneity of liver cancer and personalized therapy. Cancer Lett 2016; 379: 191-7.

[7] Trinh TL, Zhu G, Xiao X, Puszyk W, Sefah K, Wu Q, Tan W and Liu C. A synthetic aptamer-drug adduct for targeted liver cancer therapy. PLoS One 2015; 10: e0136673.

[8] Niu JX, Meng XK and Ren JJ. Studied microRNA gene expression in human hepatocellular carcinoma by microRNA microarray techniques. World J Gastroenterol 2015; 21: 12605-11.

[9] Hou X, Peng JX, Hao XY, Cai JP, Liang LJ, Zhai JM, Zhang KS, Lai JM and Yin XY. DNA methylation profiling identifies EYA4 gene as a prognostic molecular marker in hepatocellular carcinoma. Ann Surg Oncol 2014; 21: 3891-9.

[10] Okimoto G, Zeinalzadeh A, Wenska T, Loomis M, Nation JB, Fabre T, Tiirikainen M, Hernandez B, Chan O, Wong L and Kwee S. Joint analysis of multiple high-dimensional data types using sparse matrix approximations of rank-1 with applications to ovarian and liver cancer. BioData Min 2016; 9: 24.

[11] Xiong J, Xiong K and Bing Z. Clinical and RNA expression integrated signature for urothelial bladder cancer prognosis. Cancer Biomark 2018; 21: 535-546.

[12] Tang RX, Chen WJ, He RQ, Zeng JH, Liang L, Li SK, Ma J, Luo DZ and Chen G. Identification of a RNA-Seq based prognostic signature with five lncRNAs for lung squamous cell carcinoma. Oncotarget 2017; 8: 50761-50773.

[13] Gao X, Wu Y, Yu W and Li H. Identification of a seven-miRNA signature as prognostic biomarker for lung squamous cell carcinoma. Oncotarget 2016; 7: 81670-81679.

[14] van Malenstein H, Gevaert O, Libbrecht L, Daemen A, Allemeersch J, Nevens F, Van Cutsem E, Cassiman D, De Moor B, Verslype C and van Pelt J. A seven-gene set associated with chronic hypoxia of prognostic importance in hepatocellular carcinoma. Clin Cancer Res 2010; 16: 4278-88.

[15] Goldman M, Craft B, Swatloski T, Ellrott K, Cline M, Diekhans M, Ma S, Wilks C, Stuart J, Haussler D and Zhu J. The UCSC cancer genomics browser: update 2013. Nucleic Acids Res 2013; 41: D949-54.

[16] Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D and Zhu J. Exploring TCGA pan-cancer data at the UCSC cancer genomics browser. Sci Rep 2013; 3: 2652.

[17] Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, Haussler D and Zhu J. The UCSC cancer genomics browser: update 2015. Nucleic Acids Res 2015; 43: D812-7.

[18] Zhang S, Zheng R, Zeng H and Chen W. The incidence differences among sex, geographical areas and mean age of diagnosis for liver cancer in China, 1989-2008. Zhonghua Yu Fang Yi Xue Za Zhi 2014; 48: 355-60.

[19] Gagliotti C, Morsillo F, Moro ML, Masiero L, Procaccio F, Vespasiano F, Pantosti A, Monaco M, Errico G, Ricci A, Grossi P and Nanni Costa A; SInT Collaborative Study Group. Infections in liver and lung transplant recipients: a national prospective cohort. Eur J Clin Microbiol Infect Dis 2018; 37: 399-407.

[20] Tarleton HP, Chang SC, Park SL, Cai L, Ding B, He N, Hussain SK, Jiang Q, Mu LN, Rao J, Wang H, You NC, Yu SZ, Zhao JK and Zhang ZF. Genetic variation at 8q24, family history of cancer, and upper gastrointestinal cancers in a Chinese population. Fam Cancer 2014; 13: 45-56.

[21] Volk ML and Lok AS. Is family history of liver cancer a risk factor for hepatocellular carcinoma? J Hepatol 2009; 50: 247-8.

[22] Shangguan H, Tan SY and Zhang JR. Bioinformatics analysis of gene expression profiles in hepatocellular carcinoma. Eur Rev Med Pharmacol Sci 2015; 19: 2054-61.

[23] Khalid SS, Hamid S, Siddiqui AA, Qureshi A and Qureshi N. Gene profiling of early and advanced liver disease in chronic hepatitis C patients. Hepatol Int 2011; 5: 782-8.
[24] Wang YN, Shen KT, Ling JQ, Gao XD, Hou YY, Wang XF, Qin J, Sun YH and Qin XY. Prognostic analysis of combined curative resection of the stomach and liver lesions in 30 gastric cancer patients with synchronous liver metastases. BMC Surg 2012; 12: 20.
[25] Emura T, Chen YH and Chen HY. Survival prediction based on compound covariate under Cox proportional hazard models. PLoS One 2012; 7: e47627.
[26] Shi KQ, Lin Z, Chen XJ, Song M, Wang YQ, Cai YJ, Yang NB, Zheng MH, Dong JZ, Zhang L and Chen YP. Hepatocellular carcinoma associated microRNA expression signature: integrated bioinformatics analysis, experimental validation and clinical significance. Oncotarget 2015; 6: 25093-108.
[27] Di Poto C, Ferrarini A, Zhao Y, Varghese RS, Tu C, Zuo Y, Wang M, Nezami Ranjbar MR, Luo Y, Zhang C, Desai CS, Shetty K, Tadesse MG and Ressom HW. Metabolomic characterization of hepatocellular carcinoma in patients with liver cirrhosis for biomarker discovery. Cancer Epidemiol Biomarkers Prev 2017; 26: 675-683.
[28] Ha Y, Shim JH, Kim SO, Kim KM, Lim YS and Lee HC. Clinical appraisal of the recently proposed Barcelona Clinic Liver Cancer stage B subclassification by survival analysis. J Gastroenterol Hepatol 2014; 29: 787-93.
[29] Wu Y, Zang WD and Jiang W. Functional analysis of differentially expressed genes associated with glaucoma from DNA microarray data. Genet Mol Res 2014; 13: 9421-8.
[30] Imamichi T, Yang J, Huang DW, et al. Interleukin-27 induces interferon-inducible genes: analysis of gene expression profiles using Affymetrix microarray and DAVID[M]. Cytokine Protocols. Humana Press; Totowa, NJ. 2012. pp. 25-53.
[31] Liu P, Jiang W, Ren H, Zhang H and Hao J. Exploring the molecular mechanism and biomakers of liver cancer based on gene expression microarray. Pathol Oncol Res 2015; 21: 1077-83.
[32] Shangguan H, Tan SY and Zhang JR. Bioinformatics analysis of gene expression profiles in hepatocellular carcinoma. Eur Rev Med Pharmacol Sci 2015; 19: 2054-61.
[33] Mueller KM, Themanns M, Friedbichler K, Kornfeld JW, Esterbauer H, Tuckermann JP and Moriggl R. Hepatic growth hormone and glucocorticoid receptor signaling in body growth, steatosis and metabolic liver cancer development. Mol Cell Endocrinol 2012; 361: 1-11.
[34] Zerrad-Saadi A, Lambert-Blot M, Mitchell C, Bretes H, Collin de l'Hortet A, Baud V, Chereau F, Sotiropoulos A, Kopchick JJ, Liao L, Xu J, Gilgenkrantz H and Guidotti JE. GH receptor plays a major role in liver regeneration through the control of EGFR and ERK1/2 activation. Endocrinology 2011; 152: 2731-41.
[35] Bowen ME, Weninger K, Brunger AT and Chu S. Single molecule observation of liposome-bilayer fusion thermally induced by soluble N-ethyl maleimide sensitive-factor attachment protein receptors (SNAREs). Biophys J 2004; 87: 3569-84.
[36] Wang T, Li L and Hong W. SNARE proteins in membrane trafficking. Traffic 2017; 18: 767-775.
[37] Ferro-Novick S and Jahn R. Vesicle fusion from yeast to man. Nature 1994; 370: 191-3.
[38] Rothman JE. Mechanisms of intracellular protein transport. Nature 1994; 372: 55-63. [39] Yi P, Higa A, Taouji S, Bexiga MG, Marza E, Arma D, Castain C, Le Bail B, Simpson JC, Rosenbaum J, Balabaud C, Bioulac-Sage P, Blanc JF and Chevet E. Sorafenib-mediated targeting of the AAA, ATPase p97/VCP leads to disruption of the secretory pathway, endoplasmic reticulum stress, and hepatocellular cancer cell death. Mol Cancer Ther 2012; 11: 2610-20.
[39] Grigoletto A, Neaud V, Allain-Courtois N, Lestienne P and Rosenbaum J. The ATPase activity of reptin is required for its effects on tumor cell growth and viability in hepatocellular carcinoma. Mol Cancer Res 2013; 11: 133-9.
[40] Xu J, Xie R, Liu X, Wen G, Jin H, Yu Z, Jiang Y, Zhao Z, Yang Y, Ji B, Dong H and Tuo B. Expression and functional role of vacuolar H(+)-ATPase in human hepatocellular carcinoma. Carcinogenesis 2012; 33: 2432-40.
[41] Xu Z, Wang F, Fan F, Gu Y, Shan N, Meng X, Cheng S, Liu Y, Wang C, Song Y and Xu R. Quantitative proteomics reveals that the inhibition of Na(+)/K(+)-ATPase activity affects S-phase progression leading to a chromosome segregation disorder by attenuating the aurora a function in hepatocellular carcinoma cells. J Proteome Res 2015; 14: 4594-602.
[42] Wang H, Huo X, Yang XR, He J, Cheng L, Wang N, Deng X, Jin H, Wang N, Wang C, Zhao F, Fang J, Yao M, Fan J and Qin W. STAT3-mediated upregulation of lncRNA HOXD-AS1 as a ceRNA facilitates liver cancer metastasis by regulating SOX4. Mol Cancer 2017; 16: 136.
[43] Liao CG, Kong LM, Zhou P, Yang XL, Huang JG, Zhang HL and Lu N. miR-10b is overexpressed in hepatocellular carcinoma and promotes cell proliferation, migration and invasion through RhoC, uPAR and MMPs. J Transl Med 2014; 12: 234.

[44] Li Q, Ding C, Chen C, Zhang Z, Xiao H, Xie F, Lei L, Chen Y, Mao B, Jiang M, Li J, Wang D and Wang G. miR-224 promotion of cell migration and invasion by targeting Homeobox D 10 gene in human hepatocellular carcinoma. J Gastroenterol Hepatol 2014; 29: 835-42.

[45] Guo Y, Peng Y, Gao D, Zhang M, Yang W, Linghu E, Herman JG, Fuks F, Dong G and Guo M. Silencing HOXD10 by promoter region hypermethylation activates ERK signaling in hepatocellular carcinoma. Clin Epigenetics 2017; 9: 116.

[46] Qiu X, Zheng J, Guo X, Gao X, Liu H, Tu Y and Zhang Y. Reduced expression of SOCS2 and SOCS6 in hepatocellular carcinoma correlates with aggressive tumor progression and poor prognosis. Mol Cell Biochem 2013; 378: 99-106.

[47] Greenhalgh CJ, Rico-Bautista E, Lorentzon M, Thaus AL, Morgan PO, Willson TA, Zervoudakis P, Metcalf D, Street I, Nicola NA, Nash AD, Fabri LJ, Norstedt G, Ohlsson C, Flores-Morales A, Alexander WS and Hilton DJ. SOCS2 negatively regulates growth hormone action in vitro and in vivo. J Clin Invest 2005; 115: 397-406.

[48] Cui M, Sun J, Hou J, Fang T, Wang X, Ge C, Zhao F, Chen T, Xie H, Cui Y, Yao M, Li J and Li H. The suppressor of cytokine signaling 2 (SOCS2) inhibits tumor metastasis in hepatocellular carcinoma. Tumour Biol 2016; 37: 13521-13531.

[49] Qi HL, Li CS, Qian CW, Xiao YS, Yuan YF, Liu QY and Liu ZS. The long noncoding RNA, EGFR-AS1, a target of GHR, increases the expression of EGFR in hepatocellular carcinoma. Tumour Biol 2016; 37: 1079-89.