Research Article

Identification of Novel Gene Signature Associated with Cell Glycolysis to Predict Survival in Hepatocellular Carcinoma Patients

Ming Wang,1 Feng Jiang,2,3 Ke Wei,4 Erli Mao,5 Guoyong Yin,6 and Chuyan Wu5

1Department of Plastic and Burn Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210000, China
2Pediatric Department, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210000, China
3Neonatal Department, Obstetrics and Gynecology Hospital of Fudan University, Shanghai 200011, China
4Medical Service Section, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210000, China
5Department of Rehabilitation Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210000, China
6Department of Orthopedics, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210000, China

Correspondence should be addressed to Guoyong Yin; guoyong_yin@sina.com and Chuyan Wu; chuyan_w@hotmail.com

Received 22 February 2021; Accepted 22 April 2021; Published 6 May 2021

Academic Editor: Achuthan Chathrattil Raghavamenon

Copyright © 2021 Ming Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. As hepatocellular carcinoma (HCC) is a complex disease, it is hard to classify HCC with a specific biomarker. This study used data from TCGA to create a genetic signature for predicting the prognosis of HCC patients.

Methods. In a group of HCC patients (n = 424) from TCGA, mRNA profiling was carried out. To recognize gene sets that differed significantly between HCC and normal tissues, an enrichment study of genes was carried out. Cox relative hazard regression models have been used to identify genes that are significantly associated with overall survival. To test the function of a prognostic risk parameter, the following multivariate Cox regression analysis was used. The log-rank test and Kaplan–Meier survival estimates were used to test the significance of risk parameters for predictive prognoses.

Results. Eight genes have been identified as having a significant link to overall survival (PAM, NUP155, GOT2, KDELR3, PKM, NSDHL, ENO1, and SRD5A3). The 377 HCC patients were divided into eight-gene signature-based high/low-risk subgroups. The eight-gene signature's prognostic ability was unaffected by a number of factors.

Conclusion. To predict the survival of patients with HCC, an eight-gene signature associated with cellular glycolysis was then identified. The findings shed light on cellular glycolysis processes and the diagnosis of patients with low HCC prognoses.

1. Introduction

Hepatocellular carcinoma (HCC) is the most prevalent primary type of liver cancer [1, 2]. This is a heterogeneous tumor with multiple genetic and epigenetic events and is typically associated with particular risk factors such as hepatitis B or C infection, excessive consumption of alcohol, hemochromatosis, or nonalcoholic fatty liver disease induced by insulin resistance and obesity [3]. It is also the world’s fifth most prevalent cancer and the second most influential cause of cancer mortality in people. Liver cancer accounts for 70–85 percent of the overall economic burden of cancer [4, 5]. While some improvement in its clinical diagnosis and treatment has been made in recent years, HCC’s metastasis and recurrence rates after radical resection are still high. Specific diagnosis criteria and clinical targets are also desperately required for this disorder.

In the modern age of “omics,” the advent of a number of innovative techniques such as sequencing and microarray also accelerated the quest for biomarkers [6–8]. Many biomarkers of HCC have been identified, such as α-fetoprotein (AFP) and des-γ-carboxy prothrombin (DCP) [9]. Via database mining, we established thousands of biomarkers which may be correlated with tumor patient
prognosis. Since HCC is a dynamic disorder, HCC with a specific biomarker is also challenging to characterise. Studies have shown that analyzing genetic traits affecting several genes may boost prediction of prognoses [10, 11]. Specific treatment strategies can be guided by the polygenic prognostic characteristics of primary tumor biopsy. Latest research also investigated the impact of polygenic markers on HCC to determine prognosis and classify prospective patients with high-risk HCCs.

Aerobic glycolysis is one of the important characteristics of tumor, which provides survival advantage for tumor. At present, most people think that malignant tumor is not only a genetic disease but also an energy metabolic disease [10, 12]. Many glycolytic enzymes can stimulate cancer cell growth, and this “Warburg effect” is reported in various tumor forms [13, 14]. The creation of a new gene signature correlated with glycolysis may also forecast HCC. Genes were selected in this study using the gene set enrichment analysis (GSEA). GSEA is a new computing method, which can reveal more general trend of data, rather than just detect gene expression differences [15]. This approach thus strengthens the mathematical study of biological speech and biological context.

In our analysis, we collected glycolysis-related genomes from 424 HCC cases with full TCGA database mRNA expression datasets. We have verified the primary glycolysis-related mRNAs and built up an eight-gene risk signature that can predict patient prognosis accurately. Interestingly, this signature of the risk associated with glycolysis will accurately identify patients in the high-risk community who have low prognosis on multiple pathways.

2. Methods

2.1. Patients’ Clinical and mRNA Expression Data Collection. In TCGA (https://portal.gdc.cancer.gov/), we have collected clinical evidence and mRNA expression profiles from the hepatocellular cancer patients [16]. The trial included clinical data from 377 patients and age, gender, grade, stage, topography of the tumor (T), distant metastasis (M), and lymph node status (N) (Table 1).

2.2. Gene Set Enrichment Analysis. We selected five gene sets that are most closely related to glycolysis for GSEA (http://www.broadinstitute.org/gsea/index.jsp) analysis to determine whether there are significant differences in the recognized gene sets between the HCC group and normal group [17]. First, the expression level of 56753 mRNAs in the liver and neighboring tissues was examined. Finally, we determined the function of the follow-up analysis with the standardized P value (P < 0.05).

2.3. Data Processing and Calculation of Risk Parameters. The log2 transformation was used to normalize single mRNA from the expression profiles. Univariate Cox regression analysis has been used to identify genes correlated with total survival (OS) and has been exposed to multivariate Cox regression to test prognostic genes and gain coefficients.

2.4. Statistical Analysis. Kaplan–Meier survival curves have been used to measure the value of the risk parameter. Multivariate Cox analysis and stratification data analysis were performed to check whether the risk parameters were independent of clinical characteristics such as age, gender, grade, stage, tumor topography (T), distant metastasis status (M), and lymph node status (N). P < 0.05 was found statistically important. R software (v3.6.1) was used for all statistical analysis.

3. Results

3.1. Initial Gene Screening Using GSEA. From the TCGA report, we received clinical features from 377 HCC patients along with expression details for 56753 mRNAs. The expressive signatures of the glycolysis gene sets have been derived from the MSigDB database by concentrating multiple gene sets. GSEA was performed to decide if the gene sets detected differed considerably between tumor tissues and normal tissues. 3 gene sets, including the Hallmark, Reactome, and Reactome modulation of glycolysis, were significantly enriched with standardized P values <5% of the five gene sets most correlated with glycolysis (Table 2 and Figures 1 and 2).

3.2. Identification of Survival-Associating mRNAs Related to Glycolysis. First, the univariate Cox regression study was carried out with 226 genes for early screening, and 201 genes
### Table 2: Gene sets enriched in hepatocellular carcinoma (377 samples).

<table>
<thead>
<tr>
<th>Gene set</th>
<th>Size</th>
<th>ES</th>
<th>NOM P value</th>
<th>Rank at Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOCARTA_GLYCOLYSIS_PATHWAY</td>
<td>3</td>
<td>0.752</td>
<td>0.224</td>
<td>4380</td>
</tr>
<tr>
<td>HALLMARK_GLYCOLYSIS</td>
<td>199</td>
<td>0.525</td>
<td>( P \leq 0.001 )</td>
<td>11805</td>
</tr>
<tr>
<td>KEGG_GLYCOLYSIS_GLUCONEOGENESIS</td>
<td>62</td>
<td>−0.372</td>
<td>0.145</td>
<td>1079</td>
</tr>
<tr>
<td>REACTOME_GLYCOLYSIS</td>
<td>71</td>
<td>0.624</td>
<td>0.005</td>
<td>10406</td>
</tr>
<tr>
<td>REACTOME_REGULATION_OF_GLYCOLYSIS</td>
<td>12</td>
<td>0.621</td>
<td>0.037</td>
<td>8630</td>
</tr>
</tbody>
</table>

The changes in eight filtered genes were then evaluated by analyzing 377 HCC samples in the database of cBioPortal (http://cbioportal.org) [13]. The findings revealed that 35 (9.3%) of sequenced instances had changed the queried genes. The PAM gene included 3 missense mutations samples. The NUP155 gene included 3 missense mutations samples and 1 splice mutation sample. The KDELR3 gene was altered in 1.1% of cases. The PKM and SRD5A3 genes were modified in 0.3% of cases, with the NSDHL and ENO1 genes changed in 1.4% and 2.8% of cases, respectively (Figure 3(a)).

The expression differences between adjacent normal tissues and HCC tissues were also compared with 8 genes. The expression rates of the 8 genes have been greatly enhanced or reduced in HCC tissues (Figures 3(b) and 3(c)).

#### 3.3. Construction of an 8-mRNA Signature to Predict Patient Outcomes.** The forecast score model was developed on the basis of a linear combination of weighted expression rates and regression coefficients from the Cox regression multivariate analysis: risk score = 0.2193 × expression of PAM + 0.4542 × expression of NUP155 - 0.2835 × expression of GOT2 + 0.1396 × expression of KDELR3 - 0.1785 × expression of PKM + 0.3203 × expression of NSDHL + 0.1829 × expression of ENO1 + 0.3133 × expression of SRD5A3. We estimated the outcomes and graded the patients by a mean risk value into high and low categories (Figure 4(a)). The life period of every patient (in years) is shown in Figure 4(b), and the high-risk patients showed higher mortality rates than the low-risk patients. In addition, a heatmap (Figure 4(c)) was released to show the expression profiles of the 8 mRNAs, utilizing the median risk score as a cutoff for patients to use the 8-mRNA survival risk score in a low-risk or high-risk category. The ROC curve review value was 0.717 (Figure 5), which showed that the 8-mRNA signature was well adapted and unique to the metastasis and survival of HCC patients. The amount of expression of dangerous mRNA (PAM, NUP155, KDELR3, NSDHL, ENO1, and SRD5A3) in the high-risk community was higher than that of the low-risk category. In comparison, in the high-risk community, the expression level mRNA type (GOT2 and PKM) was lower than in the low-risk category.

#### 3.4. Risk Parameter Derived from 8-mRNA Signature is an Independent Prognostic Indicator.** We have contrasted the prognostic meaning of risk parameters with clinical pathological parameters through univariable and multivariate analyses (Table 1). Samples were chosen with...
well-established clinical evidence. For the 377 HCC cases, the mean age was 65. Among the 377 patients, 255 (67.6%) were male and 122 (32.4%) were female. Among 372 patients, 235 (63.2%) had grade I-II tumors and 137 (36.8%) had grade III-IV tumors. Moreover, 262 (74.2%) of the 353 HCC patients suffered from stage I-II disease and 91 (25.8%) of the remaining patients suffered from stage III-IV disease. From the above, the risk parameter and stage were determined as independent prognostic indicators because these factors demonstrated significant variations both in univariate and in multivariate analyses (Table 4). The risk parameters displayed important $P < 0.05$ (HR = 1.770) prognostic values, in particular (Figure 6).

### 3.5. Validation of Eight mRNA Markers for Survival Prediction by Kaplan–Meier Curve Analysis

Kaplan–Meier assessments of survival found that high-risk patients have a weak

<table>
<thead>
<tr>
<th>mRNA</th>
<th>Ensemble ID</th>
<th>Location</th>
<th>B (Cox)</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAM</td>
<td>ENST00000438793</td>
<td>Chromosome 5: 102,201,430–102,366,809</td>
<td>0.219257548</td>
<td>1.245151921</td>
</tr>
<tr>
<td>NUP155</td>
<td>ENST00000231498</td>
<td>Chromosome 5: 37,288,239–37,371,283</td>
<td>0.454216998</td>
<td>1.574939719</td>
</tr>
<tr>
<td>GOT2</td>
<td>ENST00000245206</td>
<td>Chromosome 16: 58,741,035–58,768,261</td>
<td>−0.283535511</td>
<td>0.753116378</td>
</tr>
<tr>
<td>KDELR3</td>
<td>ENST00000216014</td>
<td>Chromosome 22: 38,864,067–38,879,452</td>
<td>0.139601816</td>
<td>1.14981587</td>
</tr>
<tr>
<td>PKM</td>
<td>ENST00000335181</td>
<td>Chromosome 15: 72,491,373–72,523,547</td>
<td>−0.178542466</td>
<td>0.836488534</td>
</tr>
<tr>
<td>NSDHL</td>
<td>ENST00000370274</td>
<td>Chromosome X: 151,999,511–152,038,273</td>
<td>0.320316797</td>
<td>1.377564104</td>
</tr>
<tr>
<td>ENO1</td>
<td>ENST00000234590</td>
<td>Chromosome 1: 8,921,061–8,938,749</td>
<td>0.182914904</td>
<td>1.200712228</td>
</tr>
<tr>
<td>SRD5A3</td>
<td>ENST00000264228</td>
<td>Chromosome 4: 56,212,276–56,239,263</td>
<td>0.313286399</td>
<td>1.367913244</td>
</tr>
</tbody>
</table>
Figure 3: Continued.
Univariate OS regression analysis of Cox found many predictive HCC-related clinicopathologic parameters, including age, gender, grade, stage, tumor topography, distant metastasis status, and lymph node status. We then used survival figures from Kaplan–Meier to test these findings, which provided clear outcomes, with weak prognosis correlated with patients that suffered from stage III-IV cancer and with tumor topography 3-4 (Figures 7(b) and 7(c)). These results further confirmed the reliability of the analysis.

Further stratified analysis for data processing has also been performed. As shown in the K–M curve, irrespective of age, class, or grade (e.g., grade I-II or grade III-IV); the eight-mRNA signature was a reliable prognostic marker for high-risk HCC patients with poor prognosis (Figure 8). The danger parameter cannot, however, be used separately for such subgroups in view of the specific subgroups of stage III-IV, T3-4. Maybe, it is because there are relatively few normal samples in these groups. That point calls for further exploration.

4. Discussion
In recent years, studies have shown that it is not accurate to use one or several clinical features to evaluate the prognosis of tumors. Therefore, more and more research studies focus...
Figure 4: The eight-mRNA signature associated with risk parameter predicts OS in patients with endometrial cancer. (a) mRNA risk parameter distribution in each patient. (b) Survival days of EC patients in ascending order of risk parameters. (c) A heatmap of nine genes’ expression profile.
on mRNA and regard it as a biological marker of tumor progression and prognosis. For example, AFP-L3 is considered to be a specific biomarker of HCC [18]. In the early stage of hepatocarcinogenesis, the expression of squamous cell carcinoma antigen (SCCA) complexed with IgM increased, which may be an important serum biomarker for early detection of hepatocarcinoma [19, 20]. Because gene expression is easily influenced by many factors, it cannot be used as a reliable and independent prognostic indicator in many cases. In fact, HCC is a complex disease, so it is difficult to use a single biomarker to determine its nature. Therefore, the study of binding biomarkers may provide valuable reference for diagnosis and prognosis. To boost the prediction, a mathematical model consisting of genetic

![ROC curve (AUC = 0.717)](image)

**Figure 5**: Receiver operating characteristic (ROC) analysis of the sensitivity and specificity of the risk score model.

<table>
<thead>
<tr>
<th>Clinical feature</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>HR.95L</td>
</tr>
<tr>
<td>Age</td>
<td>1.009</td>
<td>0.994</td>
</tr>
<tr>
<td>Gender</td>
<td>0.812</td>
<td>0.553</td>
</tr>
<tr>
<td>Grade</td>
<td>1.143</td>
<td>0.889</td>
</tr>
<tr>
<td>Stage</td>
<td>1.622</td>
<td>1.326</td>
</tr>
<tr>
<td>RiskScore</td>
<td>1.927</td>
<td>1.587</td>
</tr>
</tbody>
</table>

**Table 4**: Univariable and multivariable analyses for each clinical feature.

![Univariable and multivariable analyses for each clinical feature](image)

**Figure 6**: Univariable and multivariable analyses for each clinical feature. (a) Univariable analysis. (b) Multivariable analysis.
The markers of several linked genes along with the predictive influence of each constituent gene is used. This model is more reliable to determine the prognosis of tumor patients than a biomarker, so it is commonly used [21, 22].

The existing high-throughput sequencing technology can extract a large number of genomic data from a single sample to determine new diagnosis, prognosis, or pharmacological biomarkers [23]. And mathematical simulations have estimated the prognosis of certain cancers with the advancement of gene marker technologies. In patients with lung adenocarcinoma, for example, a new signature to inhibit metastasis and survival was discovered by means of Cox regression and ROC study. [10]. In this study, we identified three functional glycoside gene sets that are closely related to GSEA. As above, we have selected the top-level screening feature for glucose metabolism and survival in cancer patients. The prognostic value of 8 gene combinations was determined by univariate and multivariate Cox regression analyses. Compared with other known prognostic indicators, the selected risk model may be a more targeted and powerful prognostic evaluation method, which can be used as a more effective classification tool for patients with HCC.

We used TCGA’s HCC data collection to gather genes linked to glycolysis and compare standard and HCC tissue results. Then, we selected Kaplan–Meier survival assessment, and the results showed that the prognosis of patients with low-risk parameters was relatively good. Among the eight genes, the expression of nup155 in HCC is considered to be part of the p53 regulatory network [24]. GOT2 has been shown to be involved in the energy metabolism of tumor cells. KDELR3 is considered to be one of genes that formed 11-gene-based prognostic signature of uveal melanoma [25]. PKM can promote anabolism and regulate glycolysis and

---

**Figure 7:** Kaplan–Meier survival analysis for HCC patients in TCGA dataset. (a) K–M survival curve for HCC patients with high/low risk. (b), (c) Clinical features including stage and tumor topography predict patients’ survival.
Figure 8: Continued.
promote tumorigenesis by glycolysis and control gene expression [26]. NSDHL is believed to be closely related to cholesterol metabolism. And in the study of glioma and lung cancer, ENO1 has been considered to be a possible promoter of tumor metabolism and make those tumor cells with high expression of ENO1 have growth advantage [27]. SRD5A3 is considered to be a target of prostate cancer treatment [28]. However, we did not find the relationship between PAM, metabolism, and tumor. The traditional prognosis system usually does not estimate the risk stratification and clinical results accurately. Thus, the prediction method dependent on 8-mRNA markers will help predict metastasis and prognosis of HCC in contrast with one widely used biomarker.

Figure 8: Kaplan–Meier curves for the prognostic value of risk parameter signature for the patients divided by each clinical feature. (a) Age, (b) gender, (c) grade, (d) stage I-II, (e) T1-2, (f) M0, and (g) N0.

Aerobic glycolysis is one of the most important characteristics of tumor, which provides survival advantage for tumor, also known as the “Warburg effect” [29]. This seemingly uneconomical way of energy supply is necessary for tumor cells. It not only provides energy for the growth of tumor cells but also provides raw materials for their biosynthesis. Zuo et al. believed that PGC1-α can inhibit the Warburg effect by downregulating pyruvate dehydrogenase kinase isoenzyme 1 (PDK1), so PGC1-α is considered as a potential factor to predict the prognosis and treatment target of liver cancer patients [4]. In terms of treatment, Ma et al. found that SRC-3 was highly expressed in patients with liver cancer and interacted with c-myc, the central regulator of the Warburg effect to promote its recruitment to the glycolysis gene promoter. And SRC-3 maybe is a potential target for sorafenib resistant treatment of liver cancer [30]. Three key allosteric enzymes control aerobic glycolysis: hexokinase (HK), phosphofructokinase (PFK), and pyruvate kinase (PK). Hexokinase (HK) is a key enzyme that catalyzes the first irreversible step of glycolysis and is associated with poor prognosis in cancer patients. Xu et al. found that mir-885-5p can act on 3' UTR of hexokinase 2 (HK2), significantly reducing glucose uptake and lactate production [31]. PFK is highly expressed and activated in human cancers, including
HCC, in order to generate additional energy to support tumor growth. PFK is an important potential target that can take away from cancer cells the energy and matrix needed for macromolecular synthesis and proliferation and cause normal cells to survive. The last main enzyme in glycolysis is pyruvate kinase (PK). Among the three key enzymes, PK is the most important because it controls the final conversion of phosphoenolpyruvate to pyruvate. There are four subtypes of PK (L, R, M1, and M2), and PKM2 has been found to increase significantly in hepatoma cells and played a key role in the regulation of glycolysis. It has been reported that targeting PKM2 can enhance the therapeutic effect of HCC. In general, these important glycolytic enzymes play an important role in the growth and treatment of HCC. Glycolysis can therefore be involved in the development and occurrence of HCC. We reported first a glycolysis gene marker (PAM, NUP155, KDELR3, NSDHL, ENO1, SRD5A3, GOT2, and PKM) and then proved the prognostic value for HCC.

In conclusion, eight gene risk factors associated with glycolysis are reported which can help predict survival and prognosis in HCC patients. The greater the probability factor, the bad the prediction is. This finding will allow prospective studies to discover potential HCC treatments which will provide HCC patients with further genomic targets.

Data Availability

The datasets used to support the findings of this study are available from TCGA (https://portal.gdc.cancer.gov/repository) database.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors’ Contributions

M.W and F.J contributed equally to this work. CY.W and GY.Y designed the study; KW and M.W collected the data; F.J and CY.W did the statistic analysis; F.J and ELM prepared the manuscript draft; GY.Y revised the manuscript. All the authors approved the final manuscript.

Acknowledgments

Chuyan Wu was supported by the China Scholarship Council for a 1 year study at the University of Johns’ Hopkins.

References


