FYB2 Is a Potential Prognostic Biomarker for Hepatocellular Carcinoma

Yifan Qu, Xiaozhong Shen, Xinpei Yuan and Bing Lu

Abstract: FYB2 (also known as C1orf168 or ARAP) is an adaptor protein involved in T-cell receptor (TCR)-mediated T-cell activation and adhesion. However, the correlation of FYB2 with prognosis and cancer needs further investigation. In this study, we analyzed the expression levels of FYB2 in hepatocellular carcinoma (LIHC) tumor tissues and correlated it with the pathological stages, survival outcomes, and tumor grades. We found that the expression of FYB2 was significantly downregulated in LIHC. Low FYB2 level leading to weak survival outcomes is linked with advanced tumor grades and elevated pathological stages. Cox regression analysis showed that FYB2 and AJCC-M stages can be used as independent prognostic factors for LIHC. GSEA analysis revealed that FYB2 would be notably correlated with the cellular metabolism-related pathways and particularly involved in the regulation of cancer-related pathways. Single-cell transcriptome analysis revealed that FYB2-positive cells were mainly distributed in hepatocytes, and compared with other cells, the upregulated genes of these cells were mainly enriched in metabolism-related functions. The results of the spatial transcriptome revealed that the expression of FYB2 in the adjacent area was higher than in the tumor area. These results showed that FYB2 is likely to be a new prognostic biomarker in LIHC and would help provide individual treatment decisions for LIHC patients.

Keywords: FYB2; prognostic biomarker; liver hepatocellular carcinoma

1. Introduction

Liver cancer is a type of malignant liver tumor. It can be categorized into primary and secondary types. Hepatocellular carcinoma (LIHC) is one of the main histologic subtypes of liver cancer with a high mortality rate [1–6]. The disease arises in the hepatocytes, which constitute the main liver and is the third major cause of cancer-related mortality. Epigenetic and genetic changes, aflatoxin exposure, diabetes, obesity, and smoking are the major risk factors for liver cancer [7,8]. To provide efficient initial treatment, such as local area ablation and early detection of liver cancer by monitoring methods, advance the survival rate of patients [9,10]. LIHC monitoring and diagnosis methods rely on some biomarkers, which can be objectively defined in body fluids. Most LIHC patients are generally diagnosed at advanced stages [11]; metastasis and recurrence are still the main challenges of LIHC treatment, which lead to a very poor prognosis. Therefore, it is urgent to determine new prognostic and diagnostic biomarkers for LIHC [12].

Although the function of the FYB2 gene has not been fully elucidated, it is known to encode a new type of T-cell adaptor protein. FYB2 is involved in T-cell receptor (TCR)-mediated T-cell activation and adhesion pathways. Previous research revealed that FYB2 may act as a key differentially expressed gene in breast cancer and might provide a novel target for drug discovery [13]. However, the correlation of FYB2 to prognosis and cancers awaits further investigation. Therefore, we conducted the study to check the prognosis function of FYB2 and its potential relationship with LIHC.
Here, we analyzed the expression of FYB2 in LIHC tumor tissues and associated it with survival outcomes, tumor grades, and pathological stages by different methods. The research results showed that FYB2 is likely to be a new prognostic biomarker in LIHC and would help provide a theoretical basis for the individual treatment decisions for LIHC patients and the future design of targeted therapeutic drugs.

2. Materials and Methods

To further explore the prognostic function of FYB2 and its potential relationship with LIHC, we performed bioinformatic analyses on publicly available datasets from the GEO, HRA, and TCGA database. Our bioinformatic analysis includes 50 normal and 374 tumor samples from the TCGA LIHC dataset, 52 normal and 115 tumor samples from the GEO GSE76427 dataset, 7 samples from the GSE112271 dataset, which is a 10x single cell transcriptome database containing 3 and 4 tumor samples from different regions of 2 different individuals, and 3 samples from HRA000437, which is a spatial transcriptome dataset. Based on the continuous progress of transcriptomic technology, we not only use traditional pan-oncogene analysis methods but also analyze through single-cell transcriptome and spatial transcriptome. We consider the study to be eligible for analysis based on the following criteria: (1) Studies of samples of hepatocellular carcinoma tissue. (2) Information on the technology and platform used for the study. (3) Studies with normal groups as control. Some of the applied methods are briefly described below.

2.1. GEPIA Analysis

The figures of “The gene expression profiles of all tumor samples and matched normal tissues” provided by GEPIA (http://gepia.cancer-pku.cn/index.html (accessed on 19 December 2020)) are not only used to display the distribution differences of FYB2 in different cancer types but also in normal and tumor tissues.

2.2. The Cancer Genome Atlas (TCGA)

Easy access to The Cancer Genome Atlas (TCGA) database allows for large-scale global gene expression profiling and database mining for potential correlations between genes and overall survival in a wide range of malignancies, including LIHC. It will also facilitate a better understanding of the impact of the genetic composition of tumors on clinical prognosis. In this paper, we used RNA-seq datasets in the TCGA database (50 normal samples and 374 tumor samples) to analyze the differences in FYB2 expression in normal and tumor tissues. Related datasets can be downloaded from https://portal.gdc.cancer.gov/projects/TCGA-LIHC (accessed on 19 December 2020).

2.3. Gene set Enrichment Analysis (GSEA)

We downloaded LIHC patient data from the NCBI-GEO database and analyzed them using GSEA V4.0.3 software. GSEA is applied to study the enrichment of the defined gene sets between the low-expression group and the high-expression group [14]. The samples are divided into a high group and a low group. The \( p \) value is calculated using 1000 permutations. The significant gene set was defined with \( p < 0.05 \). The data of LIHC patients are available for downloading at https://www.ncbi.nlm.nih.gov/gene/?term=FYB2 (accessed on 19 December 2020). The relevant raw data are integrated and placed at the end of the paper.

2.4. Statistical Analysis

The survival curve was drawn by using Kaplan–Meier chart. We further assessed the association between FYB2 levels and clinical features, where a Wilcoxon grade test was used between clinical variable groups, and three or more groups were tested using Kruskal–Wallis (K–W) method. The results generated in Oncomine will show the \( p \) value, multiple changes, and levels. Kaplan–Meier chart, PrognoScan, and GEPIA results are presented with log-rank test HR and P or Cox \( p \) values. Of which, GEPIA data are taken
from the website http://gepia.cancer-pku.cn/index.html (accessed on 19 December 2020). The correlation of gene expression was assessed using spearman correlation and statistical significance, and the p value of <0.05 is defined as statistically significant.

2.5. Single-Cell Transcriptome Sequencing Analysis

GSE115469 is a 10 × single-cell transcriptome dataset containing five healthy liver samples that clustered all cells in healthy livers into twenty cell populations with detailed cell annotation. We retrieved the matrix file and cell type annotation information from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115469 (accessed on 15 October 2022). We used Seurat (3.2.0) to do basic data processing and annotated cells using the author’s annotation information. By using clusterProfiler (3.14.3), we examined differentially expressed genes between these cells and other cells that had high FYB2 expression and conducted enrichment analysis on these differential genes. GSE112271 is a 10 × single-cell transcriptome database that contains three and four tumor samples coming from different regions of two different individuals. The expression matrix of GSE112271 was retrieved from https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/geo/query/acc.cgi?acc=GSE112271 (accessed on 15 October 2022). By using the MNN approach, the datasets of seven samples were combined using bachelor (1.2.4). We performed cell type annotation based on each population’s differentially expressed genes. We also looked for cells that expressed FYB2 and compared them to other cells to see which genes were different. The results were then compared to those of healthy livers.

2.6. Spatial Transcriptome Analysis

HRA000437 is a spatial transcriptome dataset containing different locations of primary liver cancer. We downloaded datasets of the spatial transcriptome of liver cancer from https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000437 (accessed on 15 October 2022). Seurat (3.2.0) was used to create objects, and the expression of FYB2 in each sample was examined.

3. Results

3.1. The mRNA Expression Level of FYB2 Is Downregulated in LIHC Tumor Tissues

At first, we used RNA-seq data in the TCGA LIHC dataset (50 normal samples and 374 tumors samples) and GEO GSE76427dataset (52 normal samples and 115 tumor samples) to analyze the difference in expression of FYB2 between normal and tumor tissues (Figure 1). Images showed the expression of FYB2 in normal tissues is significantly higher than that in tumor tissues. Next, we specifically compared FYB2 expression in 50 pairs of LIHC tumor tissues and adjacent normal tissues (Figure 2). By using the paired two-tailed t-test for the dataset, which gave p < 0.01, this paired plot supported the result of low expression of FYB2 in tumors.

![Figure 1. FYB2 expression levels in normal tissues and LIHC tumor tissues. TCGA LIHC dataset (A) and GEO LIHC GSE76427 dataset (B) were used to analyze the difference in expression of FYB2 between normal and LIHC tumor tissues.](image-url)
3.2. FYB2 Is Related to Survival Outcomes, Tumor Grades, and Pathological Stages

We further divided the patients, taking the median as the cutoff value, into two groups with low and high FYB2 levels. Survival prognosis for FYB2 was analyzed through batch survival analysis carried out with R. High expression of FYB2 is positively related to good survival outcomes, and the group with the low expression level of FYB2 has a significantly inferior prognosis compared with that of high expression ($p = 0.0042$) (Figure 3A). Furthermore, we examined the FYB2 expression in different cancer stages. We found that reduced expression of FYB2 was associated with the disease severity ($p = 0.007$) (Figure 3B). Significant differences were observed in AJCC-T stages ($p = 0.005$) (Figure 3D) and AJCC-M stages ($p = 0.015$) (Figure 3F), but significant differences were not observed in the associations of tumor grade (Figure 3C), AJCC-N stages (Figure 3E), gender (Figure 3G), and clinical variables of age (Figure 3H).
3.3. Analysis of the Correlation of FYB2 Expression with LIHC

We used Cox proportional hazard model to check the overall survival (OS) and verify if the FYB2 expression can be used as an independent predictor. We entered the factors, such as gender, age, stage, T, M, N, grade, and FYB2 (FYB2) expression, into the univariate Cox regression analysis (Table 1) to judge their influence on the OS of LIHC patients. The significant variable exhibited in univariate analysis was a further study by multivariate analysis. We also performed the forest map of Cox regression analysis (Figure 4) and used strata of tumor risk factors to evaluate the modification of the relationship between FYB2 expression and LIHC. We found that FYB2(C1orf168) and AJCC-M staging may have an impact on the OS of LIHC patients.

Table 1. The univariate analysis and multivariate analysis of the correlation of LIHC with FYB2 expression.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Univariate Analysis</th>
<th>Multivariate Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Age</td>
<td>1.0107</td>
<td>0.9956–1.0260</td>
</tr>
<tr>
<td>Gender</td>
<td>1.2446</td>
<td>0.8431–1.8374</td>
</tr>
<tr>
<td>Grade</td>
<td>1.1192</td>
<td>0.8635–1.4474</td>
</tr>
<tr>
<td>Stage</td>
<td>1.6559</td>
<td>1.3446–2.0392</td>
</tr>
<tr>
<td>T</td>
<td>1.6380</td>
<td>1.3435–1.9971</td>
</tr>
<tr>
<td>M</td>
<td>1.1878</td>
<td>0.9562–1.4755</td>
</tr>
<tr>
<td>N</td>
<td>1.0831</td>
<td>0.8685–1.3507</td>
</tr>
<tr>
<td>FYB2</td>
<td>0.9993</td>
<td>0.9987–0.9999</td>
</tr>
</tbody>
</table>

Variables: age, 10–30, 30–50, 50–60, 61–70, 71–90; gender, female vs. male; grade, G1, G2, G3, G4; stage, Stage I, II, III, IV; M, M0, M1, MX; N, N0, N1, NX; T, T1, T2, T3, T4; FYB2. HR, hazard ratio; CI, confidence interval.

Figure 4. The forest map of Cox regression analysis. Forest map describing 95% CI and hazard ratio (HR) of the correlation of FYB2 expression with LIHC. Strata were gender, age, stage, grade, T, M, N, and FYB2 (FYB2).
3.4. Gene Set Enrichment Analyses

We conducted gene set enrichment analysis (GSEA) to discover the signaling pathways related to high and low expression of FYB2 in LIHC (Figure 5). The results showed that genes involved in DNA replication, glyoxylate and dicarboxylate metabolism, hedgehog signaling pathway, Huntington’s disease, oxidative phosphorylation, Parkinson’s disease, and ribosome were enriched in the FYB2 low expression group. These results suggested that the downregulation of FYB2 was associated with a variety of cellular metabolism-related pathways as described above, as well as with diseases, and in particular with the regulation of cancer-related pathways.

![Gene set enrichment analysis (GSEA) for FYB2.](image)

Figure 5. Gene set enrichment analysis (GSEA) for FYB2.

3.5. Single-Cell Transcriptome Sequencing Revealed Differences in FYB2+ Cells between Healthy Liver and Liver Cancer

We combined the single-cell transcriptome sequencing datasets of healthy liver and liver cancer from the GEO database to further study the expression differences of FYB2 in healthy liver and liver cancer at the single-cell level, respectively (Figure 6a,b). We discovered that FYB2+ cells were mostly present in hepatocytes in the healthy liver (Figure 6b). The upregulated genes of FYB2+ cells were then analyzed using GO enrichment analysis. According to our findings, these upregulated genes were mostly discovered in functional pathways associated with metabolism, inflammatory response, and homeostasis (Figure 6c). However, we discovered that FYB2+ cells were present in malignancies in cell types other than hepatocytes (Figure 6e). We identified genes that were specifically highly expressed in FYB2+ cells in healthy liver and liver cancer, respectively, and later found that these genes were mainly different in healthy liver and liver cancer (Figure 6f).

3.6. Spatial Transcriptome Analysis Reveals Differences in FYB2 between Tumoral and Paracancerous Regions of the Liver

Spatial transcriptomics allows visualization and quantitative analysis of the transcriptome with spatial resolution in individual tissue sections. We used the spatial transcriptome datasets from the HRA database to investigate the distribution characteristics of FYB2 in tumor and non-tumor regions on a spatial basis. After evaluating three samples obtained from the area between the tumor and the adjacent tissue in the liver, in terms of visualization and quantitative analysis, we determined that FYB2 expression in the adjacent area was higher than that in the tumor area (Figure 7a–c).
3.6. Spatial Transcriptome Analysis Reveals Differences in FYB2 between Tumoral and Paracancerous Regions of the Liver

Spatial transcriptomics allows visualization and quantitative analysis of the transcriptome with spatial resolution in individual tissue sections. We used the spatial transcriptome datasets from the HRA database to investigate the distribution characteristics of FYB2 in tumor and non-tumor regions on a spatial basis. After evaluating three samples obtained from the area between the tumor and the adjacent tissue in the liver, in terms of visualization and quantitative analysis, we determined that FYB2 expression in the adjacent area was higher than that in the tumor area (Figure 7a–c).

**Figure 6.** t-SNE visualization of combined datasets from GSE115469 (a). t-SNE visualization of FYB2+ cells (b). Go enrichment analysis of upregulated genes of FYB2+ cells (c). t-SNE visualization of combined datasets from GSE112271 (d). t-SNE visualization of FYB2+ cells (e). Venn diagram showed the intersection of highly expressed genes of FYB2+ cells between healthy cancer and liver cancer (f).
4. Discussion

LIHC has a serious threat to human health around the world [15–20]. Even though immunotherapy and targeted molecular therapy have gotten great development, the OS of LIHC is still poor owing to the lack of exact early diagnosis and prognostic markers. Because of the tumor’s heterogeneity, the clinical parameters, for instance, TNM staging, cannot meet the accuracy and individual demands of prognosis prediction. The discovery of precise prognostic biomarkers and new therapeutic targets is still urgent.

In the current study, we found that FYB2 may be possible as a new prognostic biological marker in LIHC. FYB2 is an adaptor protein involved in T-cell receptor (TCR)-mediated T-cell activation and adhesion. Related studies have already suggested that FYB2 may be a key differentially expressed gene in breast cancer, which may provide a new target for future drug development.
By the use of GEPIA, we can observe that the differential distribution of FYB2 in LIHC versus normal tissues is significantly higher than that of FYB2 in other types of cancer versus normal tissues, such as Kidney Chromophobe (KICH) and Cholangiocarcinoma (CHOL). However, the role of FYB2 in tumor progression is poorly understood. We checked the expression levels of FYB2 in the TCGA LIHC dataset and GEO GSE76427 dataset and found that FYB2 expression was significantly downregulated in LIHC. The low FYB2 levels gave poor survival outcomes and were linked with upper tumor grades and advanced pathological stages. Cox regression analysis showed that FYB2 and AJCC-M stages can be used as independent prognostic factors for LIHC. GSEA analysis revealed that FYB2 would be notably correlated with the genetic materials metabolism and particularly involved in the regulation of cancer-related pathways. These results showed that C1RF168 is likely to be a new prognostic biomarker in LIHC and would help provide a theoretical basis for the future design of targeted therapeutic drugs.

Overall, our study supposed a genomic prognostic model for OS in LIHC cases. FYB2 and AJCC-M stages can be used as independent prognostic factors for LIHC. Therefore, it will contribute to the prognosis and subsequent monitoring of LIHC patients with lower cost of molecular diagnosis, for example, whole-genome sequencing. Clarifying the mechanism of FYB2 in tumorigenesis and development will increase our understanding of cancers such as LIHC, which is of great scientific and theoretical significance.

In addition, we further explored the expression of FYB2 at the single-cell level through single-cell transcriptome and spatial transcriptome. We found that both in healthy liver and liver cancer, FYB2+ cells were mainly distributed in hepatocytes, which indicated that FYB2 was closely related to the function of hepatocytes. Spatial transcriptome analysis can further compare the gene expression characteristics of FYB2 in tumor and paracancerous regions from the spatial perspective, which can reflect that the expression of FYB2 in paracancerous regions is significantly higher than that in tumor regions.

Nevertheless, our research still has a few limits. Firstly, the data in our research are mainly from GEO and TCGA databases. It is necessary to further validate the prognostic value and expression levels of FYB2 in an independent and large clinical cohort. Secondly, the potential biological mechanism of FYB2 in LITC needs to be further elucidated through functional studies. Thirdly, the combination of various prognostic genes would be more promising and valuable. The prognostic predictive models derived from multi-gene combinations have been established and verified in a range of cancers [21–26]. We should introduce multiple-gene prognostic prediction models in our further LITC research.

In summary, we found that FYB2 might be a potential biomarker of LITC. FYB2 would be notably correlated with the genetic materials metabolism and particularly involved in the regulation of cancer-related pathways. Our study supposed a genomic prognostic model for OS in LIHC cases, which may help provide theoretical basis for the individual treatment decisions for LIHC patients and the future design of targeted therapeutic drugs.

Author Contributions: Conceptualization, Y.Q.; validation, X.S.; formal analysis, Y.Q., X.S.; resources, X.Y.; writing—original draft preparation, Y.Q., X.S.; writing—review and editing, Y.Q., X.S.; visualization, Y.Q., X.S.; supervision, B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement:
1. The gene expression profiles of all tumor samples and matched normal tissues provided by GEPIA (http://gepia.cancer-pku.cn/index.html (accessed on 19 December 2020).
2. RNA-seq datasets in the TCGA database (50 normal samples and 374 tumor samples) can be downloaded from https://portal.gdc.cancer.gov/projects/TCGA-LIHC (accessed on 19 December 2020).
6. Datasets of the spatial transcriptome of liver cancer are downloaded from https://ngdc.cncb.ac.cn/gsa-human/browse/HRAR00437 (accessed on 15 October 2022).

Acknowledgments: We acknowledge that Yifan Qu and Xiaozhong Shen contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

References


