RPS3 predicts poor overall survival in HBV-related hepatocellular carcinoma patients: a data-mining with LASSO-regression algorithm

J. SHI1,2, T. ZHANG3, Z.-G. YANG2, F.-L. CHEN4, W.-S. ZHANG2

1Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China
2Department of Integrative Medicine, Shanghai Public Health Clinical Center, Fudan University, Shanghai, China
3Health Care Center, Shanghai Public Health Clinical Center, Fudan University, Shanghai, China
4Yancheng Affiliated Hospital of Nanjing University of Chinese Medicine, Yancheng, Jiangsu, China

Abstract. – OBJECTIVE: This analysis aimed to investigate the candidate biomarkers associated with overall survival (OS) in hepatocellular carcinoma (HCC) patients.

MATERIALS AND METHODS: In the GSE14520 dataset, candidate parameters were selected and included in the Cox regression and Nomogram models through bioinformatic enrichment methods and LASSO analysis, survivor functions of candidate biomarkers were also assessed.

RESULTS: Complement and coagulation cascades including 36 differential expressed genes (DEGs) and ribosome pathway including 27 DEGs were significantly enriched (both \( p < 0.05 \) and adjusted \( p < 0.05 \)). LASSO model, Cox regression and nomogram analysis indicated that RPS3, together with BCLC and TNM staging, were significantly associated with OS in HCC patients. Validated in the GEO series, TCGA and Human Protein Atlas (HPA) datasets, RPS3 mRNA and RPS3 protein were significantly upregulated in tumor tissues compared to that in nontumor tissues (all \( p < 0.05 \)). Upregulation of RPS3 has been linked to high alpha fetoprotein (AFP), advanced tumor stages and multinodular (all \( p < 0.05 \)). After adjusting AFP, tumor stage and multinodular, log rank analysis revealed that HCC patients with high RPS3 had unfavorable OS compared to those with low RPS3 (all \( p < 0.05 \)).

CONCLUSIONS: RPS3 upregulation in tumors might contribute to unfavorable OS in HCC patients.

Key Words: RPS3, Hepatocellular carcinoma, Overall survival, Ribosome.

Introduction

Liver cancer has a widespread distribution across the world, especially in Asian continent, and accounts for a serious threat to the people health1,2. As the seventh most frequently occurring cancer and the second most common cause of cancer death, the incidence of hepatocellular carcinoma (HCC) has been assumed to rise over the next 10-20 years3-5. Although advanced comprehensive approaches have been made for the treatment of HCC during the past few years, the prognosis remains poor and only a handful can benefit from existing anti-neoplastic therapies6-9. The liver cancer related mortality has been increased by more than 2% annually since 200710. Candidate biomarkers of HCC used for prognostic or predictive purposes may have vital clinical effects in the near future11. Currently, effective biomarkers, potential therapeutic targets and underlying predictors to guide treatment decisions still remain noticeably absent12. Detection of biomarkers in tumors is a direct and cost-effective adjunct, especially in monitoring disease prognosis and the selection of treatment targets in HCC patients12,13. High-throughput technical means and gene chips have become fast approaches for assessing differentially expressed genes (DEGs) and functional pathways and led to a significant increase in the availability of molecular insights involved in HCC progressiveness at multiple biological levels14-16. The accessibility of genomic sequen-
ing data from liver tumors has provided us with invaluable resources. Through integrating data from different public sources might help us to facilitate the identification of promising biomarkers or therapeutic targets.

This study identified DEGs in the GSE14520 dataset and addressed the most significant functional pathways using Gene Set Enrichment Analysis (GSEA) enrichment. The least absolute shrinkage and selection operator (LASSO) model was used to address potential candidate biomarkers for predicting overall survival (OS) in HCC patients. We hope our results could provide valuable insights into understanding of the progress and prognosis in HCC patients.

Materials and Methods

Ethics Statement

As stated by Roessler et al, all participants provided written informed consent. The informed consent and study protocol of the primary study were approved by the Institutional Review Board of Zhongshan Hospital, Fudan University. The protocol of this secondary analysis was reviewed and approved by the Ethics Committee, Shanghai Public Health Clinical Center, Fudan University.

Data Source

Microarray series and clinical dataset of GSE14520 were downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database. Two cohorts were included in the GSE14520. For gene expression profiling, tumors and paired non-tumor tissues were profiled separately using a single channel array platform. Tumor and paired non-tumor samples in cohort 1 were carried out on Affymetrix Human Genome U133A 2.0 Array platform. All samples including tumor and paired non-tumor samples of cohort 2 were processed on the Affymetrix HT Human Genome U133A Array platform.

Patients

In the GSE14520 dataset, 220 hepatitis B virus (HBV) related HCC patients were included in this analysis after excluding 27 cases without gene expression or clinical data. All liver tissue was obtained with informed consent from patients who underwent radical resection between 2002 and 2003 at the Liver Cancer Institute and Zhongshan Hospital, Fudan University.

Outcome

The primary outcome of OS was defined as the time from surgery to death from any causes.

Identification of DEGs

Raw CEL files of the microarray GSE14520 dataset were normalized by quantile method of robust multichip analysis (RMA) from R affy package. Missing gene expression data was imputed with k-Nearest Neighbor method by impute index in R program. DEGs between tumor and nontumor tissues were identified by Limma package in R program with the criterion of log2FC > 1.0 and adjusted $p$-value < 0.05. This identify framework was also addressed in other GEO series including GSE33006, GSE45436, GSE60502 and GSE84042.

Candidate gene expression between tumor and nontumor tissues were validated in The Cancer Genome Atlas (TCGA), GEO series and the Human Protein Atlas (HPA) databases. The mRNA normalized counts data of LIHC in TCGA database derived from RNAseq Htsq platform was downloaded from Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). In the Repository section, the liver and intrahepatic bile ducts were selected in the Primary Site, and TCGA-LIHC was selected in the Project Id part, transcriptome profiling and gene expression quantification were selected as Data Category and Data Type respectively. In addition, the RNA-Seq method was selected in the Experimental Strategy. Then, “Add All Files to Cart” and download the Clinical, Sample Sheet, Metadata, Manifest and Cart files in the Cart section for the following analysis.

TCGA RNAseq data contains 424 samples with 374 tumor and 50 nontumor samples. In TCGA dataset, the edgeR package in R program was used to identify gene expression levels between tumor and nontumor tissues. Protein levels with mean integral optical density (IOD) between HCC and normal samples detected by immunohistochemistry (IHC) in the HPA dataset was calculated by Image-Pro Plus version 6.0 (Media Cybernetics Inc., Rockville, MD, USA).

Enrichment Analysis

R package clusterProfiler was used to address function enrichment pathway analysis of DEGs between tumor and nontumor tissues in GSE14520 dataset. Top 20 pathways and significant pathways were presented.
significant pathways with both \( p \)-value < 0.05 and adjusted \( p \) value < 0.05 were selected for LASSO model establishment.

**The LASSO Model Establishment**

LASSO regression model was used to determine the most powerful prognostic markers for overall survival in HCC patients\(^\text{30}\). In the GSE14520 dataset, parameters including age, gender, alanine aminotransferase (ALT), cirrhosis, main tumor size, multinodular, alpha-fetoprotein (AFP), Barcelona Clinic Liver Cancer (BCLC) staging, TNM staging and Cancer of the Liver Italian Program (CLIP) staging and genes in significantly enriched pathways were included in the LASSO model. “glmnet” and “survival” packages were used for LASSO model establishment with family equals to “cox” and alpha equals to 1. The model was validated with 5-fold cross-validation. Both “lambda.1se” and “lambda.min” were used to assess the coefficient of parameters\(^\text{31}\).

**Nomogram Model Establishment**

Parameters significantly associated with OS in HCC patients in multivariate Cox model were included in the risk prediction model by nomogram with “rms” package in R program. Based on Cox proportional hazards model, “survival” package in R program was used to calculate the cumulative risk of death. To calculate the concordance index and its 95% confidence intervals (CI), “survcomp” package was used. Bootstrap method was used for repeated sampling for internal verification of the model. Calibration plot was presented for evaluating the performance of nomogram, which was also established in “rms” package in R program.

**Statistical Analysis**

Differences of variables between the individual groups were analyzed using student \( t \) test, Mann–Whitney test and Chi-square test based on data types. Parameters enrolled in the LASSO model were included in univariate and multivariate Cox regression. Results were reported as hazard ratios (HR) with 95% CI. Log rank method was used to address the survivor functions of candidate genes for OS in HCC patients. Stata software version 16.0 (STATA Corp., College Station, TX, USA) and IBM SPSS Statistics version 26.0 (SPSS Inc., Armonk, NY, USA) were used. A two-tailed \( p \) < 0.05 were considered significance for all tests.

### Results

**Functional Enrichment of DEGs**

In total, 7439 DEGs between tumor and non-tumor tissues were identified in the GSE14520 dataset. Top 20 functional pathways of these DEGs were presented in Figure 1A. Complement and coagulation cascades including 36 DEGs and ribosome pathway including 27 DEGs were significantly enriched (both \( p \) < 0.05 and adjusted \( p \) < 0.05, Figure 1B).

**Identification of Potential Candidates for OS in HCC**

In the GSE14520 dataset, 63 significant genes in complement and coagulation cascades (n = 36) and ribosome (n = 27) pathways, together with clinico-pathological characteristics including age, gender, HBV status, multinodular, cirrhosis, main tumor size, AFP, BCLC staging, TNM staging, and CLIP staging were enrolled in LASSO model (Figure 2A). After the 5-fold cross validation, parameters including RPS3, TNM staging and BCLC staging were recruited to be underlying candidates of OS in HCC patients when \( \lambda \) took the minimum value (Figure 2A). The regression coefficient plot of factors by LASSO was shown in Figure 2B.

As summarized in Table I, univariate Cox model indicated that RPS3, TNM staging and BCLC staging were underlying predictors for OS in HCC patients (all \( p \) < 0.05, Table I). After adjusting TNM staging and BCLC staging, multivariate Cox regression demonstrated that high RPS3 was significantly associated with unfavorable OS in HCC patients (HR = 1.95, 95% CI = 1.22-3.14, \( p \) = 0.006, Table I).

In addition, nomogram model was established according to the independent parameters including RPS3, TNM staging and BCLC staging (Figure 3A). According to the upper scale of each independent risk factor (-0.6 to 0.8), the corresponding score of this risk factor could be determined. The total score was obtained by adding the scores of each factor. Projecting downward from the total score, the corresponding mortality risk prediction probability value could be obtained. The concordance index of this model was 0.74 (95% CI = 0.68-0.8). Calibration curves of 1-year, 3-year and 5-year for internal verification of this nomogram with bootstrap were described in Figure 3B, 3C and 3D, respectively.
RPS3 predicts poor overall survival in HBV-related hepatocellular carcinoma patients

RPS3 Expression Between Tumors and Nontumor Tissues

RPS3 mRNA was significantly upregulated in tumor tissues compared to nontumor tissues in GSE14520 ($p < 0.001$, Figure 4A), which was validated in TCGA dataset ($p = 0.004$, Figure 4B) and Gene Expression Omnibus (GEO) series including GSE33006, GSE45436, GSE60502 and

Figure 1. Enrichment of functional pathways of differential expressed genes (DEGs) in GSE14520. Top 20 pathways (A) and significant pathways (B) of DEGs enrichment.

Figure 2. Parameter selection through LASSO regression (A) and elucidation of LASSO coefficient profiles for selected factors (B).
Table I. Univariate and multivariate Cox regression models for identifying candidates of overall survival from HCC patients#.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate</th>
<th></th>
<th>Multivariate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>p-value</td>
<td>HR (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>RPS3, high vs. low</td>
<td>2.32 (1.48-3.64)</td>
<td>&lt; 0.001</td>
<td>1.95 (1.22-3.14)</td>
<td>0.006</td>
</tr>
<tr>
<td>TNM staging</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I Reference - Reference</td>
<td>-</td>
<td>2.15 (1.24-3.73)</td>
<td>0.007</td>
<td>2.1 (1.2-3.68)</td>
</tr>
<tr>
<td>II Reference - Reference</td>
<td></td>
<td>5.09 (2.89-8.95)</td>
<td>&lt; 0.001</td>
<td>2.52 (1.11-5.73)</td>
</tr>
<tr>
<td>III Reference - Reference</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCLC staging</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 Reference - Reference</td>
<td>-</td>
<td>2.92 (0.91-9.36)</td>
<td>0.072</td>
<td>4.04 (0.98-16.64)</td>
</tr>
<tr>
<td>A Reference - Reference</td>
<td></td>
<td>6.33 (1.78-22.5)</td>
<td>0.004</td>
<td>4.42 (0.92-21.27)</td>
</tr>
<tr>
<td>B Reference - Reference</td>
<td></td>
<td>12.1 (3.58-40.89)</td>
<td>&lt; 0.001</td>
<td>8.99 (1.82-44.41)</td>
</tr>
</tbody>
</table>

#Only variables selected in LASSO model were included.

Figure 3. Competing risk model with nomogram for overall survival (OS) based on Cox regression (A) and calibration curves for 1-year (B), 3-year (C) and 5-year OS (D).
RPS3 predicts poor overall survival in HBV-related hepatocellular carcinoma patients

Figure 4. RPS3 mRNA expression between tumor and nontumor tissues in GSE14520 (A), and validation in TCGA (B) and other GEO series including GSE33006, GSE45436, GSE60502 and GSE84042 (C); RPS3 protein levels between HCC and normal samples detected by immunohistochemistry (IHC) in the Human Protein Atlas (HPA) database (D) and mean integral optical density (IOD) of these IHC samples (E).

GSE84042 (all \( p < 0.05 \), Figure 4C). In the HPA database, RPS3 protein levels of 6 HCC patients and 1 healthy individual detected by IHC were presented in Figure 4D. The means of IOD of HCC patients were significantly higher than those in normal tissues (\( p = 0.037 \), Figure 4E).

**Associations Between RPS3 and Clinico-Pathological Features in HCC Patients**

HCC patients with high RPS3 levels had higher proportions of AFP > 300 ng/ml compared to those with low RPS3 levels \[ 69 \ (62.7\%) \text{ vs. } 30 \ (27.3\%), \ p < 0.001, \text{Figure 5A}]. HCC patients with high RPS3 levels also had advanced tumor stages including BCLC staging, CLIP staging and TNM staging (all \( p < 0.05 \), Figure 5B-5D). Moreover, multinodular occurred more frequently in HCC patients with high RPS3 levels \[ 31 \ (28.2\%) \text{ vs. } 14 \ (12.7), \ p = 0.004, \text{Figure 5E}].

**Associations Between RPS3 and OS in HCC Patients**

In the GSE14520 dataset, log rank method revealed that RPS3 upregulation accounted for worse OS in HCC patients \( p = 0.002, \text{Figure 6A} \). Since RPS3 expression was associated with clinico-pathological characteristics (Figure 5), we addressed survivor function analysis adjusting these potential confounding factors. As shown in Figure 6, HCC patients with high RPS3 levels had significantly unfavorable OS compared to
those with low RPS3, after adjusting AFP, BCLC staging, CLIP staging, TNM staging and multinodular (all \( p < 0.05 \), Figure 6B-6F).

**Discussion**

Ribosomes are involved in the process of cell proliferation, growth and survival\(^{32}\). Current evidences have elucidated that inhibition of ribosome biogenesis could induce p53 stabilization and activation\(^{33,34}\). Recently, ribosome biogenesis has been linked with various human malignancies including HCC and emerged as an effective target in cancer therapy\(^{35}\). A series of compounds that inhibit ribosome biosynthesis or function have shown their toxic action on cancer chemotherapy\(^{36-38}\). Recent literatures of HCC ribosome profiling have also
RPS3 predicts poor overall survival in HBV-related hepatocellular carcinoma patients

Ribosomal protein subunit 3 (RPS3) possesses an endonuclease activity that mediates regulatory functions in DNA repair processes and plays multifunctional roles in cell apoptosis and transcriptional regulation. Knockdown of RPS3 alleviates cells injury and increases cell survival-rate from genotoxic stress after exposure to hydrogen peroxide. In addition, RPS3 serves as an essential subunit of NF-κB signaling involved in the regulation of key genes in rapid cellular activation responses. Knockdown of RPS3 impaired NF-κB mediated transcription process of selected p65 target genes. Moreover, RPS3a upregulation increases the solubility of highly aggregation-prone HBx and it could induce oncogenesis through enhancing the HBx-induced NF-κB signaling in HCC. Current evidence indicated that RPS3 functions as a vital component in regulation of tumorigenesis, immune and inflammatory responses, and cell development. As a member of ribosomal protein family, RPS3 might exert oncogenic functions synergistically with other subunits in ribosomes. Hence, oncogenic functions of RPS3, together with ribosomes signaling pathway, should be investigated experimentally and clinically in future.

Figure 6. Log rank analysis of associations between RPS3 and OS from HCC patients in GSE14520 (A), and survivor functions of RPS3 for OS adjusted for AFP (B), BCLC staging (C), CLIP staging (D), TNM staging (E) and multinodular (F) from HCC patients in GSE14520 series.

provided insightful data resource for disecting the translome shift in liver cancer, at sub-codon resolution, and the regulatory mechanisms of oncogenic signaling and HCC therapy.
RPS3 was also associated with many other human malignancies\textsuperscript{52,54}. In gastric cancer, exosomal RPS3 was essential for inducing chemoresistance of receptor cells. Exosomal RPS3, derived from cisplatin-resistant gastric cancer cell, enhanced the chemoresistance of cisplatin-sensitive gastric cancer cells through the PI3K-Akt-cofilin-1 signaling pathway\textsuperscript{55}. RPS3 knockdown significantly reduced the proliferation, survival, migration and invasion and increased apoptosis of the colon cancer cells Caco-2, which correlated with p53 elevation and lactate dehydrogenase (LDH) down-regulation\textsuperscript{56}. Similarly, the knockdown of RPS3 inhibited cell growth and induced apoptosis in breast cancer cells\textsuperscript{57}. In radiation resistant non-small cell lung cancer cells, ionizing radiation led to casein kinase 2α (CK2α)-mediated phosphorylation of RPS3, which induced dissociation of RPS3- tumor necrosis factor receptor-associated factor 2 (TRAF2) complex and NF-κB activation, resulting in significant elevation of prosurvival genes, namely, cIAP1, cIAP2, and survivin\textsuperscript{58}.

In our analysis, BCLC staging and TNM staging were also identified as predictors for HCC prognosis. Consistent with current consensus\textsuperscript{59,60}, we have addressed that BCLC staging was an unfavorable predictor for HCC prognosis previously\textsuperscript{61,62}. The predictive value of TNM staging in HCC survival has already been assessed\textsuperscript{63}. Moreover, TNM system was a better staging model for HCC compared to other applied staging systems\textsuperscript{64,65}.

This study has some limitations. The primary is that no experiments were conducted to validate the expression of RPS3 mRNA and RPS3 protein in our own biobank, and there is no experimental data to address the effects of RPS3 on hepatoma cellular functions. Secondly, this analysis was conducted at mRNA level, links between RPS3 protein and HCC prognosis was not investigated. Thirdly, this is a study based on public datasets, and none of our own follow-up data of HCC patients were available. Fourthly, the predictive values of RPS3 for OS in HCC patients were not validated in prospective cohorts.

Conclusions

In this report, we found that RPS3 was up-regulated in HCC tumors, and correlated with tumor aggressiveness. The LASSO analysis, Cox regression model, and Kaplan-Meier methods demonstrated that RPS3 was correlated with unfavorable survival in HCC patients, even after adjusting clinico-pathological characteristics including AFP, tumor staging, and multinodular. The mechanisms and clinical applications of RPS3 in HCC should be investigated in-depth in future.

Conflict of Interest
The Authors declare that they have no conflict of interests.

Acknowledgements
This study was not supported by any pharmaceutical company or government agency or grants from other sources.

Authors’ Contribution
J. Shi and T. Zhang contributed equally to this work. W.-S. Zhang and F.-L. Chen conceived and designed the study. J. Shi and T. Zhang wrote the manuscript. W.-S. Zhang, F.-L. Chen and Z.-G. Yang rewriting the manuscript. J. Shi, W.-S. Zhang, T. Zhang, and Z.-G. Yang analyzed and interpreted the data. All authors read and approved the final manuscript.

ORCID ID
J Shi: 0000-0002-9936-6939; T Zhang: 0000-0002-1735-2948; Z.-G. Yang: 0000-0002-6623-4841; F.-L. Chen: 0000-0003-2349-9812; W.-S. Zhang: 0000-0002-9081-773X.

Data Availability
Datasets of the current study are available from the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/, GSE14520, GSE33006, GSE45436, GSE60502 and GSE84042) and The Cancer Genome Atlas (https://portal.gdc.cancer.gov/). The combined datasets were available from the corresponding author (W.-S. Zhang) with reasonable request.

References
RPS3 predicts poor overall survival in HBV-related hepatocellular carcinoma patients


28) Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression


57) Ono H, Iizumi Y, Goi W, Sowa Y, Taguchi T, Sakai T. Ribosomal protein S3 regulates XIAP ex-
RPS3 predicts poor overall survival in HBV-related hepatocellular carcinoma patients


