Research Paper

High Expression of PQBP1 and Low Expression of PCK2 are Associated with Metastasis and Recurrence of Osteosarcoma and Unfavorable Survival Outcomes of the Patients

Yong Zhang¹*, Haiyue Zhao²*, Wu Xu¹, Dinghua Jiang¹, Lixin Huang¹**, Lisong Li¹**

1. Department of Orthopedic Surgery, The First Affiliated Hospital of Soochow University, No.188 Shizi Road, Suzhou 215006, China.
2. Center of Reproduction and Genetics, Suzhou Municipal Hospital, Affiliated Suzhou Hospital of Nanjing Medical University, 26 Daoqian Road, Suzhou, Jiangsu 215002, China.

* These authors contributed equally to this work

© Ivyspring International Publisher. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC) license (https://creativecommons.org/licenses/by-nc/4.0/). See http://ivyspring.com/terms for full terms and conditions.

Received: 2018.07.12; Accepted: 2019.02.23; Published: 2019.05.12

Abstract

Osteosarcoma (OS) is the most common primary bone malignancy, predominately affecting children and adolescents. Due to the introduction of chemotherapy, the 5-year survival rate of OS patients has dramatically improved to 60–70%. Unfortunately, OS patients with recurrence or metastatic disease have less than a 20% chance of long-term survival, despite aggressive therapies. In this study, we aimed to identify gene expression patterns associated with metastasis and recurrence in order to identify potential biomarkers with prognostic power. We found that high expression of polyglutamine tract-binding protein 1 (PQBP1) and low expression of phosphoenolpyruvate carboxykinase 2 (PCK2) were related to a high probability of recurrence and metastasis in OS patients and also predicted shorter recurrence-free survival (RFS) and metastasis-free survival (MFS) after adjustment for other clinical variables. Prediction models based on the combination of PQBP1 and PCK2 expression had good and robust predictive power for recurrence and metastasis. A PQBP1 and PCK2-centered protein interaction network was built, and the hypothetical regulatory path between them was identified and termed the PQBP1-SF3A2-UBA52-PCK2 axis. Gene enrichment analysis indicated that aberrations of metabolism might play an important role in recurrence and metastasis in OS patients. Accordingly, PQBP1 and PCK2 are crucial for recurrence and metastasis in OS, and these findings provide a molecular basis for the exploitation of diagnostic and therapeutic strategies for overcoming recurrence and metastasis in OS.

Key words: osteosarcoma, recurrence, metastasis, PQBP1, PCK2

Introduction

Osteosarcoma (OS) is the most common primary bone malignancy worldwide, accounting for approximately 6% of all cancers and 56% of malignant bone cancer cases in children and adolescents [1]. Although OS predominantly occurs in individuals aged 0 to 24 years old, a second peak of incidence is observed in the elderly [2]. Due to the introduction of chemotherapy and surgical removal of the tumor, the 5-year survival rate of patients with localized OS has dramatically improved over the past decades, increasing to nearly 60–70% [3]; however, no significant improvement in survival has been
reported since the 1970s [4, 5]. Unfortunately, due to local recurrence [6] or distant metastasis [7], the current treatments for OS do not always yield satisfactory clinical responses. Approximately 20% of OS patients present with pulmonary metastasis at diagnosis; furthermore, despite intensive chemotherapy and aggressive surgery, 30-40% of patients initially diagnosed without metastatic disease develop metastasis or recurrence [8, 9]. OS patients with recurrent or metastatic disease have less than a 20% chance of long-term survival, despite the use of aggressive therapies [5, 10].

Recurrence and metastasis are critical clinical episodes that have significant impacts on OS prognosis. However, it is difficult to predict which patients will develop recurrence or metastasis after therapy using only clinical indicators, such as age, gender, tumor site, and tumor stage. Although magnetic resonance imaging or computed tomography is usually applied, inflammation and recurrent/metastatic tumors are sometimes difficult to distinguish. For the past few years, with the development of high-throughput techniques, many researchers have attempted to identify molecular signatures of OS metastasis and recurrence [11-15]. However, OS is a heterogeneous tumor, as confirmed by whole genome sequencing approaches [16, 17]; therefore, the lack of consistency in the biomarkers generated from different studies remains one of the major obstacles to their clinical application. Thus, identifying potential key molecules associated with OS metastasis and recurrence that are consistent across various data sources would be of great academic and clinical significance and may lead to the development of novel prediction models and targeted therapies for OS patients with recurrent or metastatic disease, thereby improving patient prognosis.

In this study, we aimed to identify gene expression patterns associated with metastasis and recurrence in order to identify potential biomarkers with prognostic power. We analyzed three OS datasets generated using different microarray platforms and identified two genes associated with metastasis and recurrence in multiple datasets.

Methods

OS patient datasets

We searched the National Center for Biotechnology Information Gene Expression Omnibus database (GEO; https://www.ncbi.nlm.nih.gov/geo/) to identify expression microarray data sets from OS patients with clinically well-defined recurrence or metastasis information. Ultimately, a total of three datasets generated using different platforms were enrolled in our study, including 27 patients with information on metastatic events from a study by Kobayashi et al. (accession number: GSE14827, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14827) [18], 37 patients with information on recurrence events and recurrence times from a study by Kelly et al. (accession number: GSE39055, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39055) [19], and 53 patients with information on metastatic events, metastasis times, survival events, and survival times from a study by Buddingh et al. (accession number: GSE21257, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21257) [20]. Detailed clinical information for the patients in the three datasets included in our study is listed in Table 1.

Acquisition and processing of all expression profiles

The standardized gene expression profile datasets were directly downloaded from the GEO database through the Bioconductor [21] package GEOquery [22]. In the case of more than one probe set mapping to the same gene, the average value of the probe sets was used as the gene expression value [23]. In consideration of the heterogeneity across the three microarray datasets in terms of systematic measurements, a z-score transformation was performed independently in each dataset [24].

Screening and validation of metastasis- and recurrence-related genes

GSE14827 and GSE39055 were used as the discovery sets to explore gene signatures related to metastasis and recurrence in OS patients, while GSE21257 was used as an independent validation dataset. Differences in the gene expression profiles of OS patients in GSE14827 who did and did not develop metastasis were compared, and similarly, differences in the gene expression profiles of OS patients in GSE39055 who did and did not develop recurrence were analyzed. Using two-tailed t-tests, genes with a p-value < 0.05 and fold change (FC) > 1.5 were considered as differentially expressed genes (DEGs) in these two discovery datasets. Visualization of DEGs in the two discovery datasets was performed using Venn diagrams created by the VennDiagram [25] package in R and heatmap plots based on the Euclidean method created using ComplexHeatmap [26] in Bioconductor. After verification in GSE21257 using two-tailed t-tests, polyglutamine tract-binding protein 1 (PQBP1) and phosphoenolpyruvate carboxykinase 2 (PCK2) were identified as candidates for further analysis.
Table 1. Summary of the clinical and pathological characteristics of OS patients in 3 included datasets.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>GSE14827</th>
<th>GSE39055</th>
<th>GSE21257</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platform</td>
<td>hgu133plus2</td>
<td>Illumina HumanHT-12 WG-DASL V4.0 R2</td>
<td>Illumina Human-6 v2.0</td>
</tr>
<tr>
<td>Sample number</td>
<td>27</td>
<td>37</td>
<td>53</td>
</tr>
<tr>
<td>Age median (range)</td>
<td>15(8-38)</td>
<td>11(4-71)</td>
<td>16(3-79)</td>
</tr>
<tr>
<td>Metastasis/Recurrence status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no metastasis/no recurrence</td>
<td>18</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>metastasis/recurrence</td>
<td>9</td>
<td>18</td>
<td>34</td>
</tr>
<tr>
<td>Month to metastasis/recurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median (range)</td>
<td>/</td>
<td>25.5(2.7 - 151)</td>
<td>9(1-44)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>17</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>female</td>
<td>10</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Histological subtype</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chondroblastic</td>
<td>3</td>
<td>/</td>
<td>6</td>
</tr>
<tr>
<td>fibroblastic</td>
<td>1</td>
<td>/</td>
<td>5</td>
</tr>
<tr>
<td>osteoblastic</td>
<td>21</td>
<td>/</td>
<td>32</td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td>/</td>
<td>10</td>
</tr>
<tr>
<td>Chemotherapeutic response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>good</td>
<td>11</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>poor</td>
<td>16</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>femur</td>
<td>15</td>
<td>/</td>
<td>27</td>
</tr>
<tr>
<td>tibia</td>
<td>10</td>
<td>/</td>
<td>15</td>
</tr>
<tr>
<td>humerus</td>
<td>2</td>
<td>/</td>
<td>8</td>
</tr>
<tr>
<td>fibula</td>
<td>0</td>
<td>/</td>
<td>2</td>
</tr>
<tr>
<td>unknown</td>
<td>0</td>
<td>/</td>
<td>1</td>
</tr>
<tr>
<td>Tumor grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/2/3/4/unknown</td>
<td>/</td>
<td>/</td>
<td>13/16/13/5/6</td>
</tr>
<tr>
<td>Necrosis percentage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median (range)</td>
<td>/</td>
<td>65% (0.65%-0.99%)</td>
<td>/</td>
</tr>
</tbody>
</table>

Survival analysis

Based on the identification of two genes (PQBP1 and PCK2) related to metastasis and recurrence in OS patients, we next evaluated their predictive ability for metastasis and recurrence. We specified PQBP1 and PCK2 mRNA expression cutoff values in GSE14827 to define a binary expression status variable on the basis of the MinPvalue criterion from the Optimal Cutpoints [27] package in R(cutoff values: PQBP1, 0.221; PCK2, 0.771). PQBP1 and PCK2 mRNA expression levels were then classified into high (+) and low (-) expression status in the other two datasets (GSE39055 and GSE21257) according to their cutoff values established in GSE14827. The odds ratio (OR), corresponding 95% confidence intervals (95% CIs), and p-values for the correlations between metastasis/recurrence and PQBP1/PCK2 expression status, individually and in combination, were estimated using Fisher’s exact test. Kaplan-Meier survival analysis was conducted to establish survival curves using the survminer [28] package for metastasis-free survival (MFS) or recurrence-free survival (RFS) according to the univariate gene expression status as well as the combined expression status of the two genes. The statistical significance was assessed by log-rank test. In addition, we also performed Kaplan-Meier survival analysis and log-rank tests on overall survival for the purpose of checking the predictive power of the PQBP1 and PCK2 mRNA expression status for late prognosis.

Using the chi-square test via the R package coin [29], we also explored the associations between the PQBP1/PCK2 mRNA expression status and various clinical characteristics that were available in each dataset, such as age, gender, histological subtype, tumor site, and chemotherapeutic response. Univariate analysis with the Cox proportional hazards regression model was used to assess the correlations between the PQBP1/PCK2 mRNA expression status and MFS in GSE21257 or RFS in GSE39055. Multivariate analyses with the Cox proportional hazards regression model was used to test whether the PQBP1 and PCK2 expression statuses were predictors of metastasis and recurrence independent of other clinical characteristics in GSE21257 and GSE39055. Additionally, univariate and multivariate Cox regression analyses for overall survival were conducted in GSE21257. However, due to the lack of information on metastasis times, analysis of the MFS in GSE14827 could not be performed.

Construction and validation of prediction models for metastasis and recurrence

Using the caret [30] package, we constructed
prediction models for metastasis in the GSE14827 dataset based on the combination of PQBP1 and PCK2 mRNA expression values using three different machine learning algorithms: random forest, linear kernel support vector machine (SVM), and artificial neural network. These training prediction models were built through five rounds of 10-fold cross validations coupled with an internal parameter debugging process in GSE14827, and then the optimal model for each machine learning algorithm was selected. Subsequently, the three best models were selected. Subsequently, the three best models were tested for their predictive ability for metastasis in GSE21257 and for recurrence in GSE39055. Using the pROC [31] package in R, the prediction abilities of these models were assessed by receiver operating characteristic (ROC) curves and the corresponding area under the ROC curve (AUC) values.

**PQBP1 and PCK2-centered interaction network**

To explore the potential regulatory patterns between PQBP1 and PCK2, we built a PQBP1 and PCK2-centered protein interaction network derived from the Search Tool for the Retrieval of Interacting Genes (STRING) database with a combined score above 600 and k nearest neighbors = 1 [32]. We adopted the fast greedy searching community detection algorithm on this network to find functional modules. We further investigated the shortest path between PQBP1 and PCK2 in this network, and then we detected pairwise associations between the expression values of the genes encoding the proteins involved in this path in GSE14827, GSE39055, and GSE21257. Functional enrichment analysis for all genes encoding proteins in this PQBP1 and PCK2-centered interaction network was performed using hypergeometric tests. Using the complete human genome as the background, terms with an adjusted p-value < 0.05 were considered significantly enriched, and the top ten most significantly enriched terms for each annotation system were visualized as word clouds [33].

All statistical analyses and data mining procedures performed in the present study were conducted with R (version: 3.2.2) and Bioconductor software.

**Results**

**High expression of PQBP1 and low expression of PCK2 correlated with OS metastasis and recurrence**

In the present study, we used two OS patient cohorts (GSE14827 and GSE39055) with clinically well-defined metastasis or recurrence information as discovery datasets to identify gene expression alterations associated with OS metastasis and recurrence. GSE14827 included nine patients that developed metastasis and 18 metastasis-free patients, while GSE39055 consisted of 18 patients with recurrence and 19 recurrence-free patients. In GSE14827, we identified 526 DEGs, 333 of which were up-regulated and 193 of which were down-regulated in patients with metastasis compared with levels in non-metastasis patients. In GSE39055, we obtained 1057 DEGs, 690 of which were up-regulated and 367 of which were down-regulated in patients with recurrence compared with levels in recurrence-free patients. Heat maps based on DEGs derived from each discovery dataset demonstrated a clear distinction between the different groups (Figure S1A–B). The complete lists of DEGs from the analyses of GSE14827 and GSE39055 are shown in Tables S1 and S2, respectively. As shown in Figure 1A, the intersecting DEGs derived from the two datasets consisted of ten genes that were up-regulated and three that were down-regulated in patients with metastasis and recurrence (Table S3).

To demonstrate the reproducibility and robustness of these DEGs in an independent dataset, GSE21257, which was generated using a different platform, was selected as a validation cohort. Because PQBP1 was the most significantly up-regulated gene and PCK2 was the most significantly down-regulated gene in the two discovery datasets (Figure 1B–C), we concentrated on these two genes for further verification in GSE21257. Consistent with the results in the discovery datasets, PQBP1 expression was markedly elevated ($p = 0.00911$) while PCK2 expression was significantly reduced ($p = 0.03054$) in patients who developed metastasis in GSE21257 (Figure 1D).

**PQBP1 and PCK2 act as independent predictors of metastasis and recurrence**

The PQBP1 and PCK2 mRNA expression levels of patients in GSE39055 and GSE21257 were divided into “high” (+) and “low” (−) expression according to their respective cutoff values derived from GSE14827. As shown in Table 2, high expression of PQBP1 and low expression of PCK2 were significantly correlated with a high rate of metastasis in GSE14827 and GSE21257 and a high probability of recurrence in GSE39055. Furthermore, in general, more significant correlations with the incidence of OS metastasis and recurrence were found by combining the expression statuses of these two genes to form four combined expression groups: PCK2(-)PQBP1(+), PCK2(+)PQBP1(+), PCK2(+)PQBP1(-), and PCK2(-)PQBP1(+). In addition, PCK2(+)PQBP1(-) was
compared with the other three groups for a binary comparison of combined expression profiles. Kaplan-Meier curves and log-rank tests revealed that both the PQBP1(+) expression group and the PCK2(-) expression group exhibited significantly poorer RFS and MFS in GSE39055 and GSE21257, respectively (Figure 2A–B). When comparing all four combined expression profiles or the PCK2(+)PQBP1(-) profile vs. the other three combined expression profiles, we observed significantly different RFS values in GSE39055 and MFS values in GSE21257 (Figure 2C–D). These findings emphasize the ability of the PQBP1 and PCK2 mRNA expression statuses to predict metastasis and recurrence.

We further statistically assessed the relationships between the expression levels of the two genes and various clinical characteristics that were available in each dataset, such as age, gender, histological subtype, tumor site, chemotherapeutic response, tumor grade, and necrosis percentage. The only significant correlation observed was the relationship between gender and PCK2 mRNA expression status in GSE21257 (Table S4).

We next conducted a univariate Cox regression analysis to further evaluate the prognostic performance of these two genes and other clinical factors for metastasis and recurrence in each dataset. The results showed that high expression levels of PQBP1 and low expression levels of PCK2 predicted poor MFS in GSE21257 and RFS in GSE39055 (Table 3–4). Multivariate Cox regression analysis further revealed that PQBP1 and PCK2 were significantly and independently associated with the probability of metastasis and recurrence in patients with OS in GSE21257 and GSE39055, after adjusting for other clinical factors.

Figure 1. Identification of genes related to recurrence and metastasis in OS patients. (A) Venn diagrams of DEGs derived from GSE39055 and GSE14827. Ten upregulated and three downregulated were common between the two datasets. (B–D) box plots of PQBP1 and PCK2 mRNA expression levels in patients that did and did not develop recurrence/metastasis in the discovery (GSE39055, GSE14827) and validation (GSE21257) datasets. Red and blue dots represent patients with and without recurrence/metastasis, respectively.

Table 2. Expression status differences of PQBP1 and PCK2 on OS metastasis/recurrence.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>PCK2(high)</th>
<th>PQBP1(high)</th>
<th>PCK2+PQBP1</th>
<th>( p_a )</th>
<th>( p_b )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR 95%CI</td>
<td>OR 95%CI</td>
<td>OR 95%CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE14827</td>
<td>0.050 0.001–0.544</td>
<td>3.659e-03</td>
<td>5.758 0.754–55.782</td>
<td>0.072</td>
<td>0.014</td>
</tr>
<tr>
<td>GSE39055</td>
<td>0.047 0.001–0.417</td>
<td>1.098e-03</td>
<td>15.311 1.739–748.736</td>
<td>3.626e-03</td>
<td>0.013</td>
</tr>
<tr>
<td>GSE21257</td>
<td>0.196 0.046–0.758</td>
<td>0.014</td>
<td>21.618 2.796–993.201</td>
<td>2.640e-04</td>
<td>0.013</td>
</tr>
</tbody>
</table>

a. This comparison were done between PCK2(+)PQBP1(-)and PCK2(+)PQBP1(+)/PCK2(-)PQBP1(+)PCK2(-)PQBP1(+).
b. Comparisons among 4 groups such as PCK2(+)PQBP1(+)/PCK2(+)PQBP1(+)PCK2(+)PQBP1(+) and PCK2(+)PQBP1(+).
Figure 2. Prognostic significance of PQBP1 and PCK2 expression status in OS. (A) Kaplan-Meier plots of RFS with different PQBP1 or PCK2 expression statuses in GSE39055. (B) Kaplan-Meier plots of MFS with different PQBP1 or PCK2 expression statuses in GSE21257. (C) Kaplan-Meier plots of RFS and MFS with four different combined expression profiles in GSE39055 and GSE21257, respectively. (D) Kaplan-Meier plots of RFS and MFS in binary comparison of combined expression profiles in GSE39055 and GSE21257, respectively. For both PQBP1 and PCK2, (-) represents low expression status and (+) represents high expression status.

Table 3. Univariate and multivariate Cox regression analyses for RFS of OS patients in GSE39055.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate analysis</th>
<th></th>
<th></th>
<th></th>
<th>Multivariate analysis</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95%CI</td>
<td>p value</td>
<td>HR</td>
<td>95%CI</td>
<td>p value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>male vs. female</td>
<td>1.283</td>
<td>0.499–3.297</td>
<td>0.605</td>
<td>2.367</td>
<td>0.802–6.986</td>
<td>0.119</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Low vs. high</td>
<td>0.449</td>
<td>0.128–1.582</td>
<td>0.213</td>
<td>0.199</td>
<td>0.040–1.001</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Necrosis status</td>
<td>low vs. high</td>
<td>3.808</td>
<td>1.093–13.26</td>
<td>0.036</td>
<td>1.581</td>
<td>0.417–5.985</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>PCK2</td>
<td>high vs. low</td>
<td>0.096</td>
<td>0.013–0.722</td>
<td>0.023</td>
<td>0.120</td>
<td>0.015–0.985</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td>PQBP1</td>
<td>high vs. low</td>
<td>6.918</td>
<td>2.396–19.98</td>
<td>3.508e-04</td>
<td>5.934</td>
<td>1.661–21.206</td>
<td>0.006</td>
<td></td>
</tr>
</tbody>
</table>

Bold text denotes P<0.05

Table 4. Univariate and multivariate Cox regression analyses for MFS of OS patients in GSE21257.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate analysis</th>
<th></th>
<th></th>
<th></th>
<th>Multivariate analysis</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95%CI</td>
<td>p value</td>
<td>HR</td>
<td>95%CI</td>
<td>p value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>male vs. female</td>
<td>2.179</td>
<td>1.013–4.685</td>
<td>0.046</td>
<td>1.801</td>
<td>0.716–4.529</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>low vs. high</td>
<td>2.446</td>
<td>1.137–5.262</td>
<td>0.022</td>
<td>1.686</td>
<td>0.625–4.546</td>
<td>0.302</td>
<td></td>
</tr>
<tr>
<td>Tumor site</td>
<td>femur</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>fibula</td>
<td>0.467</td>
<td>0.062–3.506</td>
<td>0.459</td>
<td>0.170</td>
<td>0.015–1.935</td>
<td>0.155</td>
<td></td>
</tr>
<tr>
<td></td>
<td>humerus</td>
<td>0.805</td>
<td>0.299–2.172</td>
<td>0.669</td>
<td>0.400</td>
<td>0.113–4.410</td>
<td>0.154</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tibia</td>
<td>0.667</td>
<td>0.299–1.488</td>
<td>0.323</td>
<td>0.439</td>
<td>0.157–1.226</td>
<td>0.116</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknown</td>
<td>0.862</td>
<td>0.115–6.484</td>
<td>0.885</td>
<td>2.835</td>
<td>0.108–74.743</td>
<td>0.532</td>
<td></td>
</tr>
<tr>
<td>Histological subtype</td>
<td>chondroblastic</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>fibroblastic</td>
<td>0.296</td>
<td>0.057–1.531</td>
<td>0.146</td>
<td>0.158</td>
<td>0.012–1.998</td>
<td>0.154</td>
<td></td>
</tr>
<tr>
<td></td>
<td>osteoblastic</td>
<td>0.707</td>
<td>0.266–1.878</td>
<td>0.487</td>
<td>1.592</td>
<td>0.381–6.647</td>
<td>0.524</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>0.615</td>
<td>0.187–2.019</td>
<td>0.423</td>
<td>1.700</td>
<td>0.341–8.463</td>
<td>0.517</td>
<td></td>
</tr>
<tr>
<td>Tumor grade</td>
<td>G1</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>0.454</td>
<td>0.186–1.108</td>
<td>0.083</td>
<td>0.222</td>
<td>0.069–0.715</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>0.416</td>
<td>0.160–1.085</td>
<td>0.073</td>
<td>0.205</td>
<td>0.060–0.699</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G4</td>
<td>0.699</td>
<td>0.194–2.529</td>
<td>0.586</td>
<td>0.398</td>
<td>0.090–1.750</td>
<td>0.223</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unknown</td>
<td>0.601</td>
<td>0.190–1.903</td>
<td>0.387</td>
<td>0.219</td>
<td>0.050–0.951</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>PCK2</td>
<td>high vs. low</td>
<td>0.358</td>
<td>0.135–0.826</td>
<td>0.016</td>
<td>0.294</td>
<td>0.085–0.763</td>
<td>0.014</td>
<td></td>
</tr>
</tbody>
</table>

Bold text denotes P<0.05
We further wondered whether the expression levels of PQBP1 and PCK2 were related to the late prognosis of OS patients. To determine this, we created dot and density plots for the mRNA expression distributions of PQBP1 and PCK2 based on survival status in the three datasets. The expression distributions of the two genes were not significantly different between living and deceased cases (Figure S2A–B). We also conducted Kaplan-Meier and Cox regression analyses for overall survival based on the expression statuses of the two genes in GSE21257. High expression levels of PQBP1 predicted poor overall survival, while the expression of PCK2 was not significantly related to overall survival (Figure S2C, Table S5). Due to the lack of information on survival times in GSE14827 and GSE39055, statistical analysis for overall survival could not be conducted in these datasets.

Based on the above results, we concluded that PQBP1 and PCK2 can be employed as predictors for metastasis and recurrence in OS patients but are not suitable for the prediction of late prognosis.

**Prediction models for metastasis and recurrence based on combined PQBP1 and PCK2 expression had robust predictive power**

Using three machine learning algorithms, prediction models for metastasis based on a combination of PQBP1 and PCK2 mRNA expression values were built in GSE14827 and subsequently validated in the other two datasets. As shown in Figure 3, the AUC and 95% CI values of these three models indicated good performance for metastasis in GSE21757 and for recurrence in GSE39055 (maximum AUC: 0.7879, 95% CI: 0.6609–0.9150; GSE39055: AUC: 0.7719, 95% CI: 0.6204–0.9235). Despite some differences in performance, however, models based on the combination of PQBP1 and PCK2 from all three machine learning algorithms exhibited robust predictive power for metastasis and recurrence in OS patients. These results further indicate that the combination of the expression statuses of the two genes is able to identify OS patients with high or low risk of metastasis and recurrence.

Potential regulatory relationship between PQBP1 and PCK2

To further explore the interaction relationships between PQBP1 and PCK2, we extracted the minimal PQBP1 and PCK2-centered undirected interaction network (Figure 4A). The network contained five distinct communities with 65 nodes and 400 edges in total. PQBP1 and PCK2 belonged to different communities. We found that PCK2 was the protein most connected with other proteins in the network. Furthermore, we explored the shortest path between PQBP1 and PCK2 in this network, and we found that PQBP1 and PCK2 were connected by two other proteins, SF3A2 and UBA52. Thus, the shortest path was defined as the PQBP1-SF3A2-UBA52-PCK2 axis. The four proteins in this axis belonged to two communities, with PQBP1, SF3A2, and UBA52 in same community and PCK2 belonging to another community. UBA52 and PCK2 were hub genes (genes with an extremely high connectivity) [34] in their own communities. Then, we measured the pairwise expression correlations of the four axis genes in the three datasets (Figure 4B). The expression of UBA52 and PQBP1 was always negatively correlated with...
that of PCK2. This axis may represent the potential regulatory path between PQBP1 and PCK2.

To gain insight into the biological roles of the 65 proteins involved in the PQBP1 and PCK2-centered protein interaction network, we performed gene enrichment analyses using the signaling pathways included in the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the cellular component (GOCC), molecular function (GOMF), and biological process (GOBP) categories in the Gene Ontology (GO). The top ten most significantly enriched terms according to each annotation system were visualized as word clouds (Figure 4C). Notably, the top three most significantly enriched GOBP terms were glucose metabolic process, hexose metabolic process, and monosaccharide metabolic process. According to KEGG pathway analysis, the AMPK signaling pathway, insulin signaling pathway, and glucagon signaling pathway were the three most significantly enriched pathways. Thus, both GO and KEGG pathway enrichment analyses identified genes related to PQBP1 and PCK2 that are associated with metabolic dysfunction.

Figure 4. Potential regulatory pattern between PQBP1 and PCK2. (A) The minimal PQBP1 and PCK2-centered undirected protein interaction network from the STRING database. Nodes represent proteins, and node size is proportional to the degree of connectivity with other proteins. Edges represent interactions between proteins. Five communities were detected and are colored-coded. (B) Pairwise correlations between expression profiles of the four axis genes in GSE14827, GSE21257, and GSE39055. In the upper triangle, black dots represent pairwise expression values in all samples in each dataset, and red lines represent linear regression values. In the lower triangle, blue numbers show Pearson’s correlation coefficients between two genes. (C) Word cloud plots of top ten most significantly enriched terms among genes in the network from KEGG, GOBP, GOCC and GOMF on genes in the network. The font size and gray scale of each term were proportional to the adjusted p value.
Discussion

The results of this study showed that the gene expression patterns of PQBP1 and PCK2 were associated with metastasis and recurrence, making these genes potential biomarkers with prognostic power. PQBP1, also known as NPV38, encodes a widely expressed 38-kDa protein [35, 36] that contains several protein interaction domains: a WW domain that binds to polyglutamine tracts [36], a nuclear localization signal, and a carboxy-terminal domain that interacts with splicing factors [38-40]. Mutations in PQBP1 are reportedly associated with several X-linked mental retardation disorders, including Renpenning, Hamel, Sutherland-Haan, Golabi-Ito-Hall, and Porteous syndromes [41-45]. However, the biological function of PQBP1 in tumors remains unknown, and thus far, there has been no research into the role of PQBP1 in OS.

PCK2, also known as PEPCk2, is located on chromosome 14q11.2. The protein encoded by PCK2 functions as a rate-limiting enzyme involved in gluconeogenesis and glycolysis [46, 47]. Studies have reported potential roles for PCK2 mRNA and protein in tumors, but due to discrepancies among studies, the results remain inconclusive. For example, Vincent et al. [48] showed that PCK2 mRNA was highly expressed in the thyroid, breast, kidney, and bladder and in non-small-cell lung cancer tissue. PCK2 knockdown hampered tumor cell proliferation, reduced tumor size, and increased tumor cell death [49-51]. Conversely, another study reported that overexpression of PCK2 hindered tumorigenesis and slowed the growth of tumor-repopulating cells growth [52]. Moreover, expression of the PCK2 protein gradually decreased in cisplatin-resistant bladder cancer cells compared to levels in cisplatin-naive cells, and this decrease was dependent on the cisplatin concentration [53].

In our study, a comprehensive analysis of gene expression profiles in OS patients in the GSE14827 and GSE39055 datasets was conducted, and we identified 13 intersecting genes related to metastasis and recurrence. Among these 13 genes, PQBP1 and PCK2 were successfully validated in an independent dataset (GSE21257). High expression of PQBP1 and low expression of PCK2 were associated with a high probability of metastasis and recurrence in OS patients. Multivariate Cox regression analysis revealed that the predictive power of PQBP1 and PCK2 for RFS and MFS were independent of other clinicopathological covariates, such as gender, necrosis status, and age, in OS patients from GSE39055 and GSE21257. Furthermore, metastasis and recurrence prediction models for OS patients were built based on a combination of PQBP1 and PCK2 mRNA expression values using three machine learning algorithms in GSE14827. We found that these prediction models had robust prognostic power for metastasis and recurrence in GSE39055 and GSE21257, regardless of which machine learning algorithm was adopted. These findings indicate that PQBP1 and PCK2 may allow for the stratification of OS patients in future clinical trials. We also analyzed the PQBP1 and PCK2-centered protein network in an attempt to identify the roles of these proteins in the development of metastasis and recurrence in OS, and we extracted the shortest path between the two proteins, which reflects a possible regulatory axis between PQBP1 and PCK2. Interestingly, we noted that the most significantly enriched GO terms and KEGG pathways for genes encoding proteins in the network were related to metabolic dysfunction, suggesting that aberrations in metabolism may play an important role in OS metastasis and recurrence. Previous research has reported that PCK2 downregulation leads to a reduction in energy metabolism and may subsequently decrease susceptibility to chemotherapeutic drugs [53]. This may explain why low expression of PCK2 is associated with increased likelihoods of metastasis and recurrence in OS patients. However, the molecular mechanisms underlying the predictive roles of PQBP1 and PCK2 in OS metastasis and recurrence remain unclear, and thus the above hypothesis should be further explored and confirmed in future experimental research.

Conclusions

In this study, we newly identified PQBP1 and PCK2 as potential predictive biomarkers for metastasis and recurrence in patients with OS, providing a molecular basis for the development of targeted diagnostic and therapeutic strategies for overcoming metastasis and recurrence in OS. It is our hope that these findings will enhance clinical decision-making and strengthen disease surveillance. However, the molecular mechanisms underlying the upregulation of PQBP1 and downregulation of PCK2 in metastatic and recurrent OS patients remain unclear and should be elucidated in future studies.

Abbreviations

OS: osteosarcoma; PQBP1: polyglutamine tract-binding protein 1; PCK2: phosphoenolpyruvate carboxykinase 2; RFS: recurrence free survival; MFS: metastasis free survival; GEO: Gene Expression Omnibus database; DEGs: differentially expressed
genes; OR: odds ratio; 95% CI: 95% confidence interval; SVM: support vector machine; ROC: receiver operating characteristic; AUC: area under the ROC curve; STRING database: Search Tool for the Retrieval of Interacting Genes database; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; GOCC: Gene Ontology cellular component; GOMF: Gene Ontology molecular function; GOBP: Gene Ontology biological process; HR: hazard ratio.

**Supplementary Material**

Supplementary figures and tables. [http://www.jcancer.org/v10p2091s1.pdf](http://www.jcancer.org/v10p2091s1.pdf)

**Acknowledgements**

This work was generously supported by the National Natural Science Foundation of China (No.81601992), the Natural Science Foundation of Jiangsu Province (No.BK20160543), the Research project of Jiangsu Provincal Health Department (H201417) and the Jiangsu Provincal Clinical Orthopedic Center.

**Competing Interests**

The authors have declared that no competing interest exists.

**References**