RBP7 Regulated by EBF1 Affects Th2 Cells and the Oocyte Meiosis Pathway in Bone Metastases of Bladder Urothelial Carcinoma

Yifan Liu1,2,3,†, Mingxiang Fan4,†, Shuyuan Xian4,†, Peng Hu1, Mengyi Zhang2, Xinkun Zhang1, Hao Zhang5, Junfang Zhang4, Lingli Dai4, Mingyue Lin2, Wangli Mei4, Dan Huang4, Runzhi Huang1,4,6,7,8,*, Jie Zhang4,6,9,*, Zongqiang Huang1,*

1Department of Orthopaedics, The First Affiliated Hospital of Zhengzhou University, 450052 Zhengzhou, Henan, China
2Department of Urology, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200092, China
3Shanghai Jiao Tong University School of Medicine, 200025 Shanghai, China
4Tongji University School of Medicine, Tongji University, 200092 Shanghai, China
5Department of Orthopedics, Naval Medical Center of PLA, Second Military Medical University, 200052 Shanghai, China
6Department of Spine, Department of Orthopedics, Tongji Hospital affiliated to Tongji University School of Medicine, 200065 Shanghai, China
7Department of Burn Surgery, The First Affiliated Hospital of Naval Medical University, 200433, Shanghai, China
8Research Unit of Key Techniques for Treatment of Burns and Combined Burns and Trauma Injury, Chinese Academy of Medical Sciences, 100007, Beijing, China
9Shanghai First Maternity and Infant Hospital, Tongji University School of Medicine, 201204 Shanghai, China
*Correspondence: runzhhuang2022@163.com (Runzhi Huang); jiezhang@tongji.edu.cn (Jie Zhang); gzhuangzq@163.com (Zongqiang Huang)
†These authors contributed equally.

1. Introduction

Bladder cancer has the highest incidence among urologic malignancies in China and is the sixth most common cancer globally [1,2]. One third of patients might have aggressive progression later, and 70% of patients who undergo surgical intervention will still experience metastasis [3]. The pathological type of most bladder cancer patients is urothelial carcinoma, and patients with the same pathological features usually have a considerable difference in prognostic outcome [4]. Its morbidity has risen from 2000 to 2011, and patients with metastasis were reported to have a higher mortality rate [1]. After radical cystectomy, nearly 50% of patients will still develop varying degrees of progression. Around 5–10% of patients have already metastasized at the time of diagnosis [5]. Bone metastasis occurs frequently in bladder urothelial carcinoma (BLCA), which has a relative incidence of 40% and only 6–9 months of median survival from diagnosis [6,7]. Thus, there is still an
unmet clinical need for further research to reveal the potential mechanism of bone metastasis in BLCA.

High throughput sequencing data analysis has been widely used for the identification of significant molecular and cellular biomarkers in the metastasis and prognosis of BLCA [8–12]. However, little previous research has focused on immune genes and relative networks in the bone metastasis of BLCA.

In this study, we identified differentially expressed immune genes (DEIGs) and TFs in patients with BLCA. To obtain valuable prognostic biomarkers, univariate Cox regression was applied. Based on the DEIGs, we constructed a proportional hazards model and validated it to have good prognosis-predictive performance. We employed CIBERSORT and Single Sample Gene Set Enrichment Analysis (ssGSEA) method to quantify tumor-infiltrating immune cells/pathways, which were then co-analyzed with the DEIGs. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was carried out employing the in silico tools Gene Set Variation Analysis (GSVA) and Gene Set Enrichment Analysis (GSEA). Subsequently, KEGG signaling pathways with statistically significant correlation relationships were selected. Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq), Chromatin Immunoprecipitation sequencing (ChIP-seq), single-cell RNA sequencing (scRNA-seq), and multiple online databases were utilized for external validation. Taken together, we proposed a scientific hypothesis for bone metastasis in BLCA.

2. Materials and Methods

2.1 Data Sources

RNA sequencing profiles of 408 BLCA samples were obtained from The Cancer Genome Atlas (TCGA) database (https://www.cancer.gov/tcga), and 209 BLCA samples with complete demographic and prognostic characteristics, especially bone metastasis information integrity, were eventually selected. Among the 209 BLCA samples, 11 had bone metastasis. Additionally, information on cancer-related transcription factors (TFs) was accessed through the Cistrome database (http://cistrome.org) [13]. In addition, 2498 immune-related genes and their gene signatures were retrieved from the ImmPort database (https://immport.org) [14] and the Molecular Signatures Database (MSigDB) v7.0 (https://www.gsea-msigdb.org) [15], respectively. What is more, the single-cell RNA sequencing profiles of the tumor sample, as well as CD4 and CD8 samples of BLCA patients were downloaded from the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/) (accession no. GSE164046, no. GSE149652).

2.2 Differential Gene Expression and Univariate Cox Regression Analyses

Differentially expressed genes (DEGs) were identified between bone-metastatic and non-metastatic samples with the cutoff value of false discovery rate (FDR) <0.05 and $|\log_2(\text{fold change})| > 1$ through the edgeR package. And DEIGs were subsequently recognized by matching DEGs with the 2498 immune-related genes downloaded from the ImmPort database. Gene expression values were pre-processed by z-score before fold change and p value were calculated. In addition, Gene Oncology (GO) and KEGG enrichment analysis was conducted. Subsequently, univariate Cox proportional hazards model was constructed to identify DEIGs significantly correlated with overall survival. Disease-free survival analysis of these identified DEIGs was also obtained from the Gene Expression Profiling Interactive Analysis (GEPIA) database (http://geopia.cancer-pku.cn/) [16].

2.3 Construction of a Proportional Hazards Model

Using univariate and multivariate Cox regression analyses, DEIGs with statistical significance were included in the proportional hazards model. Risk score for prognosis was calculated according to the following formula:

$$\text{Risk score}_i = \sum_{a=1}^{n} \beta a \times (\text{expression level of gene } a)$$

In the formula, “i” stands for the number of the patient, “n” stands for the number of significant DEIGs included in the model, and “$\beta a$” stands for the regression coefficient of the DEIGs. The median cut point of the risk scores was used as the cutoff value for high and low risk grouping. Model predictions were evaluated by plotting the Kaplan-Meier curve. Predictive performance was quantified using area under the curve (AUC). In order to further estimate the independent prognostic value of risk score, the univariate and multivariate Cox regression analyses were utilized. Moreover, 196 BLCA patients cohort from the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) database (https://ocg.cancer.gov/programs/target) were extracted for further validation. Differential enrichment of tumor infiltrating immune cells and immune pathways in BLCA was evaluated using the CIBERSORT and ssGSEA methods.

2.4 Differential TFs Expression and Correlation Analyses

Differentially expressed TFs were identified by the edgeR package. FDR was set to be less than 0.05 while $|\log_2(\text{fold change})| > 1$. Then, correlation analysis was performed to co-analyze DEIGs and TFs, illustrating their regulatory relationships with correlation coefficient $>0.300$ and $p < 0.001$, in which case if the variables conformed to normal distribution and homogeneity of variance, Pearson correlation analysis was conducted, or else Spearman correlation analysis was used.

2.5 Immune Cells/Pathways and KEGG Pathways Analyses

Differential enrichment of tumor infiltrating immune cells/pathways in BLCA was evaluated using the CIBER-
Table 1. Baseline information of 209 BLCA samples.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total patients (N = 209)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>67.14 ± 10.14</td>
</tr>
<tr>
<td>Median (Range)</td>
<td>67 (60–75)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>160 (76.56%)</td>
</tr>
<tr>
<td>Female</td>
<td>49 (23.44%)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>39 (18.66%)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>10 (4.78%)</td>
</tr>
<tr>
<td>White</td>
<td>148 (70.81%)</td>
</tr>
<tr>
<td>Not Reported</td>
<td>12 (5.74%)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>3 (1.44%)</td>
</tr>
<tr>
<td>II</td>
<td>78 (37.32%)</td>
</tr>
<tr>
<td>III</td>
<td>70 (33.49%)</td>
</tr>
<tr>
<td>IV</td>
<td>58 (27.75%)</td>
</tr>
<tr>
<td>M</td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>198 (94.74%)</td>
</tr>
<tr>
<td>M1</td>
<td>11 (5.26%)</td>
</tr>
<tr>
<td>N</td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>133 (63.64%)</td>
</tr>
<tr>
<td>N1</td>
<td>21 (10.05%)</td>
</tr>
<tr>
<td>N2</td>
<td>29 (13.88%)</td>
</tr>
<tr>
<td>N3</td>
<td>3 (1.44%)</td>
</tr>
<tr>
<td>NX</td>
<td>22 (10.53%)</td>
</tr>
</tbody>
</table>

Abbreviation: SD, Standard deviation; M, bone metastasis; N, lymph node metastasis.

SORT and ssGSEA tools. The correlation between key DEIGs and immune cells/pathways was evaluated by Pearson/Spearman correlation coefficient. In addition, GSVA was applied to score KEGG pathways in each sample. To explore the potential KEGG pathways correlated with prognosis and metastasis in BLCA patients, Pearson/Spearman correlation analysis and univariate Cox regression were conducted. Moreover, the Venn diagram was plotted using the VennDiagram package [17].

2.6 Construction of the Regulatory Network

Focusing on the immune genes, TFs and downstream immune cells/pathways and KEGG pathways were identified which were involved in the bone metastasis in BLCA. Therefore, we constructed a regulatory network to further elucidate the mechanism of bone metastasis in BLCA by exploiting the Cytoscape package [18].

2.7 ATAC-seq, ChIP-seq and scRNA-seq Validation

To further validate the transcriptional regulation pattern between EBF1 and RBP7, data related to EBF1 acquired from chromatin immunoprecipitation sequencing (ChIP-seq) of in vitro human cell lines were downloaded from Gene Expression Omnibus (GEO) (GEO assession: samples GSM1958038 and GSM1958039 from GSE75503, GSM935375 from GSE31477, and GSM803386 from GSE32465) [19,20]. ENCODE Transcription Factor Targets [21] and JASPAR Predicted Transcription Factor Targets [22] algorithms were then applied to re-predict the transcriptional regulation pattern between EBF1 and RBP7. Additionally, ATAC-seq data of BLCA samples were obtained from the TGCA project of chromatin accessibility landscape of primary human cancers (https://gdc.cancer.gov/about-data/publications/ATACseq-AWG), which were then utilized to explore the chromatin accessibility in specific locations of EBF1 and RBP7. What is more, the scRNA-seq profiles of BLCA samples from GEO were obtained to successively go through quality control, normalization, cell cycle assignment, feature selection, principal component analysis (PCA), Uniform Manifold Approximation and Projection (UMAP)/t-distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction and visualization, unsupervised clustering, differential expression analysis, and cell communication analysis.

2.8 Online Databases Validation

Multiple online databases were used to access the external validation of our scientific hypothesis. CellMarkers (http://biocc.hrbmu.edu.cn/CellMarker/) [23] and PathCards (https://pathcards.genecards.org/) [24] were applied to find biomarkers related to immune genes/pathways and KEGG pathways. What is more, UALCAN (http://ualcan...
Fig. 2. Differential gene expression analysis and univariate Cox proportional hazards model. (A) Heatmap demonstrated DEGs between bone-metastatic and non-metastatic BLCA samples. (B) Volcano plot showed over-expressed and under-expressed DEGs with $p < 0.05$. (C) Heatmap demonstrated DEIGs between bone-metastatic and non-metastatic BLCA samples. (D) Volcano plot showed over-expressed and under-expressed DEIGs with $p < 0.05$. (E) Forest plot illustrated 31 DEIGs correlated with overall survival by univariate Cox proportional hazards model. Abbreviation: DEGs, differentially expressed genes; DEIGs, differentially expressed immune genes.
which were defined as DEIGs and displayed in a heatmap. 1090 DEGs were found in the list of 2498 immune genes, Fig. 1A,B. In addition, GO and KEGG enrichment analyses, visualized as bubble plots, were performed to uncover that these DEGs actively participated in pathways such as axonogenesis, metal ion transmembrane transporter activity, and neuroactive ligand-receptor interaction which might play a role in patients’ survival and immune function. The expression of the 31 DEIGs were displayed in survival plots (Supplementary Fig. 2) and were distinguished by Cox univariate and multivariate analyses. All biosignature and statistical analyses were implemented using R software (Version 3.6.1; Institute for Statistics and Mathematics, Vienna, Austria; https://www.r-project.org). A significance level of a two-sided p value less than 0.05 was adopted.

3. Results

3.1 Identification of DEIGs

The analysis procedure of this study was demonstrated in Fig. 1. Clinical information (Table 1) and RNA-seq data of 209 BLCA samples were accessed through TCGA database. Between bone-metastatic and non-metastatic samples, 1090 DEGs were identified (Fig. 2A,B). Additionally, GO and KEGG enrichment analyses, visualized as bubble plots, were performed to uncover that these DEGs actively participated in pathways such as axonogenesis, metal ion transmembrane transporter activity, and neuroactive ligand-receptor interaction which might play a role in the bone metastasis of BLCA patients (Supplementary Fig. 1A,B). In addition, using differential analysis, 86 of 1090 DEGs were found in the list of 2498 immune genes, which were defined as DEIGs and displayed in a heatmap plot (Fig. 2C) and a volcano plot (Fig. 2D). Moreover, in univariate Cox regression analysis, 31 DEIGs were exhibited to have a significant association with overall survival, in which 29 risk factors and 2 protective factors were identified in a forest plot (Fig. 2E). What is more, the differences of disease free survival between low and high expression of the 31 DEIGs were displayed in survival plots (Supplementary Fig. 2).

3.2 Construction of a Multivariate Prognostic Model

Based on the 31 prognosis-related DEIGs, we constructed a multivariate prognostic prediction model. The receiver operating characteristic (ROC) curve of the model was exhibited in Fig. 3A, in which the AUC was 0.793, indicating a good predictive value. Sub-group survival analysis was conducted according to the median risk score, displaying a significant difference (p = 1.04 \times 10^{-7}) between the low risk and high risk groups (Fig. 3B). The clinical features of risk score were represented by a risk curve (Fig. 3C) and a scatter plot (Fig. 3D). The expression levels of the 31 DEIGs between low-risk and high-risk groups were demonstrated by a heatmap (Fig. 3E). Together with four clinical features (age, gender, M and stage), the risk score in our model proved to be an independent prognostic predictor by univariate (HR, 1.011; 95% CI, 1.005–1.017; p < 0.001) and multivariate (HR, 1.009; 95% CI, 1.004–1.015; p < 0.001) Cox regression analyses (Fig. 4A,B). Moreover, significant differences of censor (p < 0.001), grade (p < 0.05), T (p < 0.05), N (p < 0.001), and stage (p < 0.001) between low risk and high risk groups in a 196 BLCA patients cohort from the TARGET database (Fig. 4C). Additionally, the distribution of different immune cell components in low risk and high risk groups was demonstrated in Supplementary Fig. 3A. What is more, the comparison of different immune cell fractions (Supplementary Fig. 3B) and ssGSEA scores (Supplementary Fig. 3C) between low risk and high risk group were displayed, which revealed that the risk score was correlated with expression of CD8 T cells (p < 0.001), T cells follicular helper (p < 0.05), and macrophages M0 (p < 0.01), along with immune functions of NK cells (p < 0.001) and Th2 cells (p < 0.001).

3.3 Regulatory Relationship between TFs and DEIGs

8 differentially expressed TFs (|log2FC| > 1 and FDR < 0.05) were illustrated by a heatmap (Fig. 5A) and a volcano plot (Fig. 5B). Additionally, a venn diagram was utilized to identify 5 DEIGs which were KLRK1, IL9R, CNTFR, RBP7, and TNFSF4 related to metastasis, stage, and TF co-expression simultaneously (Fig. 5C). Aiming to identify the regulatory relationships between the identified TFs and key DEIGs, Pearson correlation analysis was utilized and 30 regulatory relationships were identified with |correlation coefficient| > 0.300 and p value < 0.001. Eventually, EBF1 positively regulating RBP7 was selected as the most potential regulatory relationship (R = 0.677, p value < 0.001). In addition, the significant differences of RBP7 expression between metastatic and non-metastatic BLCA samples (p = 0.018), as well as among stage i, ii, iii, and iv of BLCA samples (p = 0.018) were demonstrated (Fig. 5D,E).

3.4 Correlation Analysis of DEIGs, Immune Cells/Pathways and KEGG Pathways

For the purpose of exploring the potential signaling pathways related to bone metastasis and BLCA prognosis, CIBERSORT, ssGSEA, and GSVA algorithms were utilized. The correlation between expression of immune cells/pathways and RBP7 expression was analyzed by Pearson’s correlation (Supplementary Fig. 4A). Moreover, 5 immune cells/pathways were identified to be statistically significant (Supplementary Fig. 4B), including macrophages M0 (R = 0.212, p value = 0.007), Type 2 helper T (Th2) cells (R = 0.23, p value < 0.001), parainflammation (R = –0.201, p value = 0.003), neutrophils (R
**Fig. 3. Construction of a multivariate prognostic model.** (A) ROC curve revealed a good prediction reliability (AUC = 0.793). (B) Kaplan-Meier survival curve showed a significant difference ($p = 1.04 \times 10^{-7}$) between low risk and high risk groups. (C) Risk curve displayed the distribution of patients with increasing risk scores in the low risk and high risk groups. (D) Scatter plot represented the distribution of survival time and survival status of low risk and high risk groups. (E) Heatmap demonstrated differential expression of the 31 DEIGs between low risk and high risk groups. Abbreviations: DEIGs, differentially expressed immune genes.
Fig. 4. Independent prognostic analysis. (A) Univariate Cox regression analysis with age, gender, M, stage, and risk score. (B) Multivariate Cox regression analysis with age, gender, M, stage, and risk score. (C) Heatmap visualized differences of censor, age, gender, grade, TNM stage, clinical stage, and bone metastasis status between low risk and high risk groups in 196 TARGET patients (*: \( p < 0.05 \); **: \( p < 0.001 \)).

\( R = -0.179, \ p \text{ value} = 0.009 \) and MHC class I \( (R = -0.148, \ p \text{ value} = 0.032) \). What is more, 9 prognosis-related KEGG pathways were characterized to be significantly correlated with RBP7 expression (Supplementary Fig. 5A,B).

3.5 Identification of Potential KEGG Pathways

Using the GSEA, 23 metastasis-related KEGG pathways were identified. Thus, as illustrated in the Venn plot (Fig. 6A), the oocyte meiosis pathway was identified to be the key KEGG pathway both significant in GSEA and GSVA co-expression (Fig. 6B). What is more, the correlation relationship between the oocyte meiosis pathway and RBP7 \( (R = 0.14, \ p = 0.042) \) was shown in Fig. 6C. Besides, GSEA result of the oocyte meiosis was demonstrated by the enrichment plot in Fig. 6D.

3.6 Construction of the Regulation Network

The regulatory network based on TFs, RBP7, immune genes/pathways and KEGG pathways was constructed in Fig. 7A. Taken together, we proposed that the EBF1-RBP7 axis (representing TF-immune gene relationship) played an important role in bone metastasis of BLCA through Th2 cells and the oocyte meiosis pathway.

3.7 ATAC-seq Validation

ATAC-seq analysis was performed to explore the correlation between EBF1 and RBP7. Open chromatin loci on different chromosomes were displayed in Supplementary Fig. 6A. Additionally, the venn pie and upset plot showed the distribution and intersection of different pick types such asgenic, intergenic, intron, and exon in Supplementary Fig. 6B. Moreover, the distribution of binding loci rela-
Fig. 5. Identification of differentially expressed TFs and co-expression analysis. (A) Heatmap revealed differentially expressed TFs between metastatic and non-metastatic BLCA samples. (B) Volcano plot showed over-expressed and under-expressed TFs with \( p < 0.05 \) and \(|\log_2(\text{fold change})| > 1\). (C) Venn diagram presented the overlapped number of DEIGs related to bone metastasis, stage and co-expression. (D) Bee-swarm plot illustrated the differential RBP7 expression \( (p = 0.018) \) between metastatic and non-metastatic BLCA samples. (E) Bee-swarm plot displayed the differential RBP7 expression \( (p = 0.018) \) among stages i, ii, iii, and iv of BLCA samples. Abbreviations: TFs, transcription factors; DEIGs, differential expressed immune genes.

tive to TSS was visualized in Supplementary Fig. 6C. Additionally, the correlation between EBF1 and RBP7 was analyzed, which turned out that the expression of EBF1 was positively correlated with RBP7 \( (R = 0.71, p < 0.001) \) (Fig. 7B). What is more, multiple binding peaks in BLCA samples at promoters of EBF1 and RBP7, as well as various regulatory elements binding areas in the introns and in introns of neighboring genes were identified, which indicated that these regions might function as potential regulatory elements on upstream of EBF1 and RBP7 sequences (Fig. 7C,D).
Fig. 6. Co-expression and correlation analysis of the KEGG oocyte meiosis pathway. (A) Venn plot for the KEGG pathway selected using GSVA and GSEA algorithms. (B) GSEA analysis for the oocyte meiosis pathway. (C) RBP7 expression was positively correlated with the oocyte meiosis pathway. (D) GSEA analysis showed a positive correlation between the oocyte meiosis pathway and RBP7 expression. Abbreviation: KEGG, Kyoto Encyclopedia of Genes and Genomes; GSVA, Gene-Set Variation Analysis; GSEA, Gene-Set Enrichment Analysis.

3.8 ChIP-seq Validation

To further illustrate the co-expression patterns between EBF1 and RBP7, ENCODE and JASPAR algorithms were utilized. In EBF1 ChIP-seq data of activated GM12878, LCL and Mutul cell line (homo sapiens), binding peaks were all found in RBP7 sequences (Fig. 8A), and 2 motifs of EBF1 were identified (Fig. 8B,C). Consequently, the transcriptional regulation pattern between EBF1 and RBP7 could be determined.

3.9 scRNA-seq Validation

Furthermore, in the scRNA-seq analysis of BLCA tumor cell lines, 12 seurat clusters and bulk labels were shown in t-SNE dimension reduction analysis (Fig. 9A). Additionally, the top 5 marker genes of each cluster was demonstrated in a heatmap (Fig. 9B). In addition, up- and down-regulated marker genes of each cluster was revealed in a point plot (Fig. 9C). More importantly, the distribution of 5 marker genes expression including EBF1, RBP7, CD24, MKI67, and CD44 was illustrated in Fig. 9D, which indicated that EBF1 and PBP7 had a co-expression relationship in clusters 0, 1, 2, 3 and 4, and they might also be co-expressed with tumor stemness and proliferation markers CD24, MKI67, and CD44. What’s more, the distribution of different stages in a cell cycle and relevant scores were visualized in Fig. 9E, which suggested that BLCA cells with
high EBF1 and RBP7 expression had a relative high fraction of stages G2M and S cells, including clusters 2, 3 and 4. Last but not least, the cell communication among different clusters presented that clusters 1, 2 and 4 had the strongest interaction with other clusters (Fig. 9F). Moreover, in the scRNA-seq analysis of BLCA CD4 and CD8 T cell lines (Supplementary Fig. 7), a co-expression relationship between EBF1 and RBP7 was also identified.

3.10 External Validation with Multiple Online Databases

In order to access the external validation of our scientific hypothesis, multiple online databases were employed. Three biomarkers of Th2 cells (IL1RL1, GATA3, and IL33) and the top five genes related to the oocyte meiosis pathway (ADCY3, ADCY5, ADCY9, MAPK3, and IGF1) were presented using the CellMarker and PathCards databases, respectively. All the external validation results are summarized in Table 2.

In UALCAN database validation (Supplementary Figs. 8,9), high ADCY3 (Supplementary Fig. 8F) and MAPK3 (Supplementary Fig. 8I) expression levels were characterized as being significantly correlated with worse BLCA patients’ prognosis. Additionally, expression levels of EBF1 (Supplementary Fig. 9A),
Fig. 8. The results of ChIP-seq validation. In EBF1 ChIP-seq data of activated GM12878, LCL and Mutul cell lines (homo sapiens), binding peaks were all found in RBP7 sequences (A). Two EBF1-binding motifs (B,C).
Fig. 9. Single-cell RNA sequencing of BLCA tumor cell lines. (A) t-SNE dimension reduction analysis showed 12 seurat clusters and bulk labels in BLCA tumor cell lines. (B) Top 5 marker genes of each cluster in a heatmap. (C) Marker genes of each cluster in a point plot. (D) The distribution of 5 marker genes including EBF1 and RB7. (E) UMAP dimension reduction analysis showing the distribution of different stages in a cell cycle as G1, G2M and S, along with G2M.Score and S.Score in violin plot. (F) Ligand-receptors plot among 12 clusters.
RBP7 (Supplementary Fig. 9B), IL33 (Supplementary Fig. 9E), ADCY5 (Supplementary Fig. 9G), ADCY9 (Supplementary Fig. 9H), and IGFl (Supplementary Fig. 9J) were identified to be elevated in normal tissues compared to BLCA primary tumor tissues, while GATA3 expression was escalated in BLCA primary tumor tissues (Supplementary Fig. 9D).

In GEPIA database validation (Supplementary Figs. 10A–H), it was discovered that elevated EBF1, RBP7, IL1RL1, IL33, ADCY3, ADCY5, ADCY9, MAPK3, and IGFl (Supplementary Fig. 10A–H) expression levels were associated with unfavorable prognoses in BLCA patients, while over-expression of GATA3 (Supplementary Fig. 10I) indicated better overall survival. In addition, ADCY3 (Supplementary Fig. 11A), IGFl (Supplementary Fig. 11B), and EBF1 (Supplementary Fig. 11G) expression were recognized to be positively correlated with RBP7 expression, while GATA3 (Supplementary Fig. 11I) expression was negatively correlated with RBP7 expression.

In LinkedOmics database validation (Supplementary Figs. 12A,13), it was shown that escalated expression of ADCY3 (Supplementary Fig. 12A), IGFl (Supplementary Fig. 12B), ADCY9 (Supplementary Fig. 12E), MAPK3 (Supplementary Fig. 12H), and RBP7 (Supplementary Fig. 12J) were significantly correlated with poor BLCA patients’ prognosis, while over-expressed GATA3 (Supplementary Fig. 12I) suggested good prognosis. Moreover, ADCY3, IGFl, ADCY5, IL1RL1, ADCY9, IL33, EBF1, MAPK3, and GATA3 expression (Supplementary Fig. 13A–H) were illustrated to be positively correlated with RBP7 expression, while GATA3 (Supplementary Fig. 13I) expression was negatively correlated with RBP7 expression.

In TISIDB database validation (Supplementary Fig. 14), up-regulation of ADCY3 (Supplementary Fig. 14A), IGFl (Supplementary Fig. 14B), ADCY9 (Supplementary Fig. 14E), and MAPK3 (Supplementary Fig. 14H) were significantly linked to worse BLCA patients’ prognosis, while up-regulated GATA3 (Supplementary Fig. 14I) expression was significantly related to better prognosis of BLCA patients.

In cBioportal database validation (Supplementary Fig. 15), ADCY3, IGFl, ADCY5, IL1RL1, ADCY9, IL33, EBF1, and MAPK3 (Supplementary Fig. 15A–H) expression were positively correlated with RBP7 expression, while expression of GATA3 (Supplementary Fig. 15I) was negatively associated with expression of RBP7.

In UCSC Xena database validation (Supplementary Fig. 16), down-regulation of ADCY3, IGFl, ADCY5, IL1RL1, ADCY9, IL33, and EBF1 (Supplementary Fig. 16A–G) were significantly correlated with better BLCA patients’ prognosis, while down-regulated GATA3 (Supplementary Fig. 16I) expression was significantly correlated with better BLCA patients’ prognosis.

In OncoLnc database validation (Supplementary Fig. 17), lower expression of ADCY3, IGFl, ADCY5, IL1RL1, ADCY9, IL33, EBF1, and RBP7 (Supplementary Fig. 17A–J) were revealed to be significantly correlated with favorable prognosis of BLCA patients, while down-regulated GATA3 (Supplementary Fig. 17I) expression was significantly correlated with favorable prognosis of BLCA patients.

In PROGgeneV2 database validation (Supplementary Figs. 18A,19), low-expressed IGFl (Supplementary Fig. 18B), ADCY9 (Supplementary Fig. 18E) and RBP7 (Supplementary Fig. 18J) were demonstrated to be significantly associated with better BLCA patients’ prognosis. What is more, down-regulated combined gene expression of EBF1, RBP7, IL1RL1, GATA3, IL33, ADCY3, ADCY5, ADCY9, MAPK3, and IGFl was significantly correlated with favorable prognosis of BLCA patients (Supplementary Fig. 19).

4. Discussion

BLCA is a common cancer globally, requiring further investigation since patients with metastatic lesions still lack good prognosis and effective treatments. Around 25% of new patients are diagnosed with either muscle invasiveness or metastasis [32]. Locally advanced and metastatic cases usually respond poorly to current treatments and have poorer survival outcomes [33,34]. High-throughput technologies have been widely used to explore the molecular mechanisms of BLCA. Several biomarkers with different clinical utilities have been identified, and some of them are already capable of capturing prominent tumors and their molecular biology characteristics [35]. Immune genes have been identified as prognostic biomarkers in BLCA patients, which provide new opportunities to explore the bone-metastatic mechanisms in BLCA [36].

In this study, we identified 31 DEIGs with prognostic value in bone metastatic BLCA samples. Based on the 31 prognosis-related DEIGs, we developed and validated a proportional hazards model with a high effectiveness in Kaplan-Meier survival analysis (p < 0.01) and ROC curve (AUC = 0.793). The Pearson/Spearman correlation analysis was applied between DEIGs and TFs, in which EBF1 (TF) positively regulated RBP7 (DEIG) represented the most significantly correlated relationship.

As a member of early B-cell factors (EBFs) family, EBF1 was reported to play a pivotal role in the development and differentiation of B cell [37]. Genetic alteration in EBF1 was found in many B-lineage acute leukemia patients [38]. Recent research have revealed that several clinical features of acute lymphoblastic leukemia, such as pathogenesis, are related with the mutation of EBF1 [39]. Additionally, genomic loss of EBF1 has been reported in solid tumors such as breast cancer and pancreatic cancer [39,40]. In this study, we found that EBF1 was low-expressed in BLCA patients, and the difference in expression levels be-
<table>
<thead>
<tr>
<th></th>
<th>EBF1</th>
<th>RBP7</th>
<th>IL1RL1</th>
<th>GATA3</th>
<th>IL33</th>
<th>ADCY3</th>
<th>ADCY5</th>
<th>ADCY9</th>
<th>MAPK3</th>
<th>IGF1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UALCAN</strong></td>
<td>OS: ( p = 0.300 )</td>
<td>OS: ( p = 0.490 )</td>
<td>OS: ( p = 0.130 )</td>
<td>OS: ( p = 0.096 )</td>
<td>OS: ( p = 0.240 )</td>
<td>OS: ( p = 0.021 )</td>
<td>OS: ( p = 0.920 )</td>
<td>OS: ( p = 0.062 )</td>
<td>OS: ( p = 0.047 )</td>
<td>OS: ( p = 0.092 )</td>
</tr>
<tr>
<td>Expression: ( p &lt; 0.001 )</td>
<td>Expression: ( p = 0.004 )</td>
<td>Expression: ( p = 0.889 )</td>
<td>Expression: ( p = 0.029 )</td>
<td>Expression: ( p = 0.001 )</td>
<td>Expression: ( p = 0.466 )</td>
<td>Expression: ( p = 0.001 )</td>
<td>Expression: ( p = 0.001 )</td>
<td>Expression: ( p &lt; 0.001 )</td>
<td>Expression: ( p = 0.618 )</td>
<td>Expression: ( p = 0.001 )</td>
</tr>
<tr>
<td><strong>GEPIA</strong></td>
<td>OS: ( p = 0.014 )</td>
<td>OS: ( p = 0.047 )</td>
<td>OS: ( p = 0.019 )</td>
<td>OS: ( p = 0.012 )</td>
<td>OS: ( p = 0.026 )</td>
<td>OS: ( p = 0.027 )</td>
<td>OS: ( p = 0.006 )</td>
<td>OS: ( p = 0.037 )</td>
<td>OS: ( p = 0.006 )</td>
<td>OS: ( p = 0.003 )</td>
</tr>
<tr>
<td>Correlation: ( R = 0.580 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R &gt; -4.20 \times 10^{-4} ), ( p = 0.99 )</td>
<td>Correlation: ( R = -0.150 ), ( p = 0.002 )</td>
<td>Correlation: ( R = 0.017 ), ( p = 0.730 )</td>
<td>Correlation: ( R = 0.130 ), ( p = 0.008 )</td>
<td>Correlation: ( R = 0.071 ), ( p = 0.140 )</td>
<td>Correlation: ( R = 0.054 ), ( p = 0.270 )</td>
<td>Correlation: ( R = 0.097 ), ( p = 0.046 )</td>
<td>Correlation: ( R = 0.340 ), ( p &lt; 0.001 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LinkedOmics</strong></td>
<td>OS: ( p = 0.060 )</td>
<td>OS: ( p = 0.008 )</td>
<td>OS: ( p = 0.432 )</td>
<td>OS: ( p = 0.005 )</td>
<td>OS: ( p = 0.224 )</td>
<td>OS: ( p = 0.008 )</td>
<td>OS: ( p = 0.066 )</td>
<td>OS: ( p = 0.004 )</td>
<td>OS: ( p = 0.004 )</td>
<td>OS: ( p = 0.002 )</td>
</tr>
<tr>
<td>Correlation: ( R = 0.546 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.108 ), ( p = 0.030 )</td>
<td>Correlation: ( R = -0.170 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.137 ), ( p = 0.006 )</td>
<td>Correlation: ( R = 0.224 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.302 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.265 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.221 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.421 ), ( p &lt; 0.001 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TISIDB</strong></td>
<td>OS: ( p = 0.111 )</td>
<td>OS: ( p = 0.086 )</td>
<td>OS: ( p = 0.173 )</td>
<td>OS: ( p = 0.026 )</td>
<td>OS: ( p = 0.720 )</td>
<td>OS: ( p = 0.006 )</td>
<td>OS: ( p = 0.071 )</td>
<td>OS: ( p = 0.028 )</td>
<td>OS: ( p = 0.017 )</td>
<td>OS: ( p = 0.012 )</td>
</tr>
<tr>
<td>cBioportal</td>
<td>Correlation: ( R = 0.547 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.110 ), ( p = 0.026 )</td>
<td>Correlation: ( R = -0.172 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.138 ), ( p = 0.005 )</td>
<td>Correlation: ( R = 0.224 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.301 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.266 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.222 ), ( p &lt; 0.001 )</td>
<td>Correlation: ( R = 0.421 ), ( p &lt; 0.001 )</td>
<td></td>
</tr>
<tr>
<td>UCSC Xena</td>
<td>OS: ( p = 0.007 )</td>
<td>OS: ( p = 0.086 )</td>
<td>OS: ( p &lt; 0.001 )</td>
<td>OS: ( p = 0.013 )</td>
<td>OS: ( p = 0.004 )</td>
<td>OS: ( p = 0.004 )</td>
<td>OS: ( p = 0.019 )</td>
<td>OS: ( p = 0.020 )</td>
<td>OS: ( p = 0.173 )</td>
<td>OS: ( p = 0.003 )</td>
</tr>
<tr>
<td>OncoLnc</td>
<td>OS: ( p = 0.014 )</td>
<td>OS: ( p = 0.031 )</td>
<td>OS: ( p = 0.115 )</td>
<td>OS: ( p = 0.023 )</td>
<td>OS: ( p = 0.017 )</td>
<td>OS: ( p = 0.038 )</td>
<td>OS: ( p = 0.020 )</td>
<td>OS: ( p = 0.032 )</td>
<td>OS: ( p = 0.043 )</td>
<td>OS: ( p &lt; 0.001 )</td>
</tr>
<tr>
<td>PROGgeneV2</td>
<td>OS: ( p = 0.070 )</td>
<td>OS: ( p = 0.034 )</td>
<td>OS: ( p = 0.068 )</td>
<td>OS: ( p = 0.138 )</td>
<td>OS: ( p = 0.156 )</td>
<td>OS: ( p = 0.460 )</td>
<td>OS: ( p = 0.092 )</td>
<td>OS: ( p = 0.007 )</td>
<td>OS: ( p = 0.125 )</td>
<td>OS: ( p = 0.006 )</td>
</tr>
</tbody>
</table>
between normal and tumor tissue was statistically significant. Besides, high-expressed EBF1 in BLCA predicts a worse survival outlook. It is reasonable to believe that the dysfunction of EBF1 can clarify the mechanism of tumorigenesis and bone metastasis in BLCA to some extent.

It has been documented that gene-based signatures related to immunology have significant prognostic value in patients diagnosed with papillary thyroid cancer and non-squamous non–small cell lung cancer [41,42]. The retinol binding protein 7 (RBP7), is a member of the cellular retinol-binding protein (CRBP) family, whose members are required for vitamin A stability and metabolism [43,44]. More importantly, RBP7 has been reported by several recent studies to be a part of immune-related gene signatures to indicate clinical significance and predict prognosis of BLCA patients [36,45,46]. What is more, it was revealed that RBP7 was a significant biomarker whose high expression indicated poor prognosis of colon cancer patients, as well as being related to tumor invasion and epithelial mesenchymal transition [47]. Additionally, RBP7 was also discovered to be an immune-related gene signature in gastric cancer [48]. In this study, we used bioinformatics analysis and identified that RBP7 was statistically correlated with stage, bone metastasis, and overall survival in BLCA patients, which may provide us with a better understanding of bone metastasis in BLCA. Although no previous study has reported the relationship between EBF1 and RBP7, Pearson correlation analysis demonstrated that they are significantly correlated with each other (R = 0.677, p value < 0.001). What is more, the transcriptional regulation pattern between EBF1 and RBP7 was further confirmed by ATAC-seq and ChIP-esq analysis. Last but not least, utilizing scRNA-seq analysis in BLCA tumor cell lines and T cell lines, EBF1 and RBP7 were revealed to be co-expressed and related to tumor stemness and proliferation.

An increase in bladder cancer metastasis has been confirmed both in humans and mice by previous study [36]. Additionally, Th2 cells can also function as a promoter factor in tumor growth [49]. In addition, aberrant high levels of cytokines were found within the immune microenvironment mediated by Th2 cells in several cancer types [50]. Moreover, one of the cytokines, IL-10, was correlated with the low efficacy of immunotherapy of Bacillus Calmette Guérin (BCG) in bladder cancer [51]. Employing CIBERSORT, ssGSEA and Pearson/Spearman correlation analysis, Th2 cells were identified to be the immune cells which were actively involved in bone metastasis in BLCA. What is more, the biomarkers of Th2 cells, including IL1RL1, GATA3, and IL33 were also tested for the prognostic values by multiple online databases.

For further exploration of the mechanism of bone metastasis in BLCA, we performed GSVA and GSEA to find metastasis-related KEGG pathways, in which the oocyte meiosis pathway was identified. The oocyte meiosis pathway was also significantly correlated with RBP7 expression using Pearson/Spearman correlation analysis. Previous studies have demonstrated that the oocyte meiosis pathway might be related to BLCA development and bone metastasis [52,53]. Additionally, multiple online databases validated that biomarkers of the oocyte meiosis pathway, including ADCY3, ADCY5, ADCY9, MAPK3 and IGF1, were significantly correlated with bone metastasis and prognosis in BLCA patients. However, only in silico analysis was performed in this study, further validation of our hypothesis demands supplementary experiments. Unfortunately, for lack of corresponding clinical samples and with limited experimental conditions, we couldn’t perform supplementary experiments to further validate our hypothesis immediately. Nevertheless, multi-omics data analysis was utilized to make up for it. More importantly, we are now working with several clinical departments to collect corresponding clinical samples, and it is believed that relevant experiments for the validation of our findings will soon be performed in our follow-up research.

5. Conclusions

Taken together, we applied an integrated bioinformatics analysis to identify DEIGs and related pathways involved in bone metastases of BLCA and improved our understanding of the molecular mechanisms. We proposed that the EBF1-RBP7 relationship has potential importance in bone metastatic BLCA through Th2 cells and the oocyte meiosis pathway.

Availability of Data and Materials

The datasets generated and/or analysed during the current study are available in the TCGA portal (https://portal.gdc.cancer.gov).

Author Contributions


Ethics Approval and Consent to Participate

The study was approved by the Ethics Committee of First Affiliated Hospital of Zhengzhou University (No. 2021-KY-190e).

Acknowledgment

We appreciate the The Cancer Genome Atlas (TCGA) team for providing an open access resource for our study.
Funding

This study was supported in part by the Shanghai Rising-Star Program (Sailing Special Program) (No. 23YF1458400); the Shanghai Municipal Health Commission (No.201940306); Henan medical science and technology research project (No. 201602031); Key project of provincial and ministerial co-construction of Henan Medical Science and Technology (No. SBGJ202002031). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/j.fbl2808189.

References

induced transgelin promotes bladder cancer metastasis by regulating epithelial-mesenchymal transition and invadopodia formation. EBioMedicine. 2019; 47: 208–220.


