A Transcription Factor Signature Predicts Prognosis of Patients with Adrenocortical Carcinoma

Jianyu Zhao  
China-Japan Union Hospital of Jilin University

Bo Liu  
China-Japan Union Hospital of Jilin University

Xiaoping Li (xiaopingli@jlu.edu.cn)  
Jilin University First Hospital  https://orcid.org/0000-0001-6463-7922

Research

Keywords: Adrenocortical carcinoma, TCGA, GEO, transcription factor, prognosis

Posted Date: February 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-178087/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Adrenocortical carcinoma (ACC) is a rare endocrine cancer that manifests as abdominal masses and excessive steroid hormone levels. Transcription factors (TFs) deregulation is found to be involved in adrenocortical tumorigenesis and cancer progression. This study aimed to construct a TF-based prognostic signature for prediction of survival of ACC patients.

**Methods:** The gene expression profile for ACC patients were downloaded from TCGA and GEO datasets. The univariate Cox analysis was applied to identify survival-related TFs and the LASSO Cox regression was conducted to construct the TF signature. The multivariate analysis was used to reveal the independent prognostic factors.

**Results:** We identified a 13-TF prognostic signature comprised of CREB3L3, NR0B1, CENPA, FOXM1, E2F2, MYBL2, HOXC11, ZIC2, ZNF282, DNMT1, TCF3, ELK4, and KLF6 using the univariate Cox analysis and LASSO Cox regression. The risk score based on the TF-signature could classify patients into low- and high-risk group. Kaplan-Meier analyses showed that patients in the high-risk group had significantly shorter overall survival compared to the low-risk patients. ROC curves showed that the prognostic signature predicted the overall survival of ACC patients with good sensitivity and specificity. Furthermore, the TF-risk score was an independent prognostic factor.

**Conclusion:** Taken together, we identified a 13-TF prognostic marker to predict overall survival in ACC patients.

1 Background

Adrenocortical carcinoma (ACC) is a rare endocrine cancer with an annual incidence of 0.7-2.0 cases per million. It usually affects adults aged around 40–50 years and children younger than 10 years. Clinical manifestations of ACC include abdominal masses and elevated steroid hormones, and result in overall poor outcomes with five-year survival ranging from 32–45%. Therefore, it is essential to identify prognostic markers of ACC in order to screen for patients at high risk.

Transcription factors (TFs) are regulatory proteins that bind to the promoter sequences of genes and decrease or increase their transcription, and thus control cell differentiation, proliferation and death. Not surprisingly, the genes encoding TFs are often aberrantly expressed in human cancers and developmental disorders. For instance, p53 and c-Myc mutations are correlated with poor clinical outcomes in cancer patients. In addition, overexpression of the forkhead box transcription factor FoxP3 is an independent prognostic factor for the overall survival of patients with ovarian cancer, while E2F1 is related to adverse prognosis in patients with non-small cell lung carcinomas. In recent years, TF-related gene expression signatures, such as that of p53 and STAT3, have been identified in several cancers. Snail is overexpressed in numerous ACC patients and associated with decreased survival. In addition, TGF-β pathway components including GATA-6 and SF-1 are also correlated with poor outcomes in ACC patients. Therefore, it is worth investigating the correlation between TFs and ACC prognosis, and constructing a prognostic TF signature for predicting patient survival.

The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets have enabled researchers to correlate clinical outcomes with transcriptomic profiles. To this end, we systematically analyzed the gene expression data of ACC datasets using univariate and multivariate Cox regression models. Based on the survival analysis, we then developed a 13-TF prognostic signature and validated this model in an independent microarray data set from GEO.
2 Material And Methods

2.1 Data collection and preprocessing

A total of 1639 human TFs were identified from a previously published study. TCGA ACC level 3 RNAseqv2 data (RSEM_genes_normalized file) and corresponding clinical information were downloaded from TCGA database. A total of 79 patients with ACC were included after excluding those lacking complete clinical and survival data. The microarray dataset GSE19776 based on the GPL570 platform was downloaded from GEO. The gene expression profiling of 22 patients with ACC were performed using [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. The clinical data included age, gender and tumor stage. Firstly, we obtained 1554 overlapped TFs between TFs in literature and TCGA dataset, and 1508 overlapped TFs between TFs in literature and GEO dataset. A total of 118 TFs were excluded from TCGA dataset by removing lower expression genes. Next, we acquired the common TFs between TCGA and GEO datasets.

2.2 Survival analysis

After obtained common TFs between TCGA and GEO datasets, univariate Cox survival analysis was performed using Cox proportional hazards regression model. Only TFs with Cox P < 0.001 and Log Rank P < 0.001 on univariate analysis were incorporated into the Lasso Cox regression analysis. Kaplan-Meier method was used to analyze the correlation between overall survival (OS) and TF expression, and the OS of different patient groups were compared using the Log-Rank test. The “survival” package in R software was used for survival analysis and the time-dependent ROC (Receiver Operating Characteristic) curve analysis was performed using the “survival ROC” package.

2.3 Construction of ACC prognostic signature

LASSO Cox regression model was widely used for high-dimensional predictor identification. In the present study, OS-associated TFs used to to select the significant TFs associated with the OS of ACC patients according to the coefficient value. These factors were incorporated in the multivariate Cox regression model to construct the ACC prognostic signature. The risk score for each TF-encoding gene was calculated as follows:

\[ \text{Risk score} = \sum_{i=1}^{n} \text{exp}_i \times \beta_i \]

where n is the number of selected genes, exp_i is the expression level of gene i and \( \beta_i \) is the coefficient of gene i.

2.4 Gene set enrichment analysis (GSEA)

GSEA was performed to analyze the significance of the 13 TFs constituting the prognostic signature using GSEA v2.0.12 (http://www.broadinstitute.org/gsea) by computing the enrichment score for each gene set. The distributions of these TFs against the rank-ordered gene ontology (GO) hallmarks were characterized using GSEA with the default settings. Positive and negative normalized scores indicated enrichment in the high-risk and low-risk groups respectively.

3 Results

3.1 Construction of the TF prognostic signature
The scheme for developing the TF signature is outlined in Fig. 1. After initially identifying 1639 TFs by literature search, the low expressing genes were removed and 1304 common TFs were screened from TCGA and GEO datasets. Univariate regression analysis showed that 23 TFs were correlated with OS (Cox-P < 0.001 and Log-Rank P < 0.001), of which 13 were identified by the Lasso regression analysis, and the risk score was calculated by multivariate cox analysis.

The λ value was selected in the LASSO Cox regression analysis when the median of the sum of squared residuals was the smallest (Fig. 2A). The following survival-related TFs with non-zero coefficients were then screened: CREB3L3 (cAMP-responsive element-binding protein 3-like 3), NR0B1 (nuclear receptor subfamily 0, group B, member 1), CENPA (centromere protein-A), FOXM1 (Forkhead Box M1), E2F2 (E2F transcription factor 2), MYBL2 (v-myb myeloblastosis viral oncogene homolog (avian)-like 2), HOXC11 (homeobox C11), ZIC2 (Zic family member 2), ZNF282 (zinc finger protein 282), DNMT1 (DNA methyltransferase 1), TCF3 (transcription factor 3), ELK4 (ETS transcription factor ELK4) and KLF6 (Krüppel-like factor 6) (Fig. 2B). Only CREB3L3 and NR0B1 were negatively correlated with the remaining TFs (Fig. 2C), while CENPA, FOXM1 and E2F2 displayed a strong correlation (Table 1). As shown in Fig. 3, high expression of CREB3L3 (HR = 0.663, Cox P = 5e-05, Log-Rank P = 1.97e-07) and NR0B1 (HR = 0.799, Cox P = 6.93e-05, Log-Rank P = 3.18e-06) were associated with good prognosis, and that of other TFs with poor prognosis (HR > 1, P < 0.001).

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Stable ID</th>
<th>Gene type</th>
<th>Chr position(start-end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENPA</td>
<td>ENSG00000115163</td>
<td>Protein coding</td>
<td>2(26764289–26801067)</td>
</tr>
<tr>
<td>CREB3L3</td>
<td>ENSG00000060566</td>
<td>Protein coding</td>
<td>19(4153631–4173054)</td>
</tr>
<tr>
<td>DNMT1</td>
<td>ENSG00000130816</td>
<td>Protein coding</td>
<td>19(10133345–10231286)</td>
</tr>
<tr>
<td>E2F2</td>
<td>ENSG0000007968</td>
<td>Protein coding</td>
<td>1(23506438–23531233)</td>
</tr>
<tr>
<td>ELK4</td>
<td>ENSG00000158711</td>
<td>Protein coding</td>
<td>1(205597556–205631962)</td>
</tr>
<tr>
<td>FOXM1</td>
<td>ENSG00000111206</td>
<td>Protein coding</td>
<td>12(2857681–2877155)</td>
</tr>
<tr>
<td>HOXC11</td>
<td>ENSG00000123388</td>
<td>Protein coding</td>
<td>12(53973126–53977643)</td>
</tr>
<tr>
<td>KLF6</td>
<td>ENSG00000067082</td>
<td>Protein coding</td>
<td>10(3775996–3785281)</td>
</tr>
<tr>
<td>MYBL2</td>
<td>ENSG00000101057</td>
<td>Protein coding</td>
<td>20(43667019–43716495)</td>
</tr>
<tr>
<td>NR0B1</td>
<td>ENSG00000169297</td>
<td>Protein coding</td>
<td>X(30304206–30309598)</td>
</tr>
<tr>
<td>TCF3</td>
<td>ENSG00000071564</td>
<td>Protein coding</td>
<td>19(1609290–1652615)</td>
</tr>
<tr>
<td>ZIC2</td>
<td>ENSG00000043355</td>
<td>Protein coding</td>
<td>13(99981784–99986765)</td>
</tr>
<tr>
<td>ZNF282</td>
<td>ENSG00000170265</td>
<td>Protein coding</td>
<td>7(149195546–149226238)</td>
</tr>
</tbody>
</table>

### 3.2 TF signature risk score can predict prognosis of ACC patients

The clinical relevance of these TFs was further assessed by multivariate Cox regression analysis, and the risk score based on their expression levels and coefficients was calculated. The 13-TF risk score classified the patients...
from TCGA training set into the high- and low-risk groups (Fig. 4A). Expect for CREB3L3 and NR0B1, all TFs were overexpressed in high-risk group (Fig. 4B). Furthermore, ACC patients in the high-risk group had significantly shorter survival compared to the low-risk patients (HR = 16.95 (5.02–57.2); Cox $P = 5.11 \times 10^{-6}$; Log Rank $P = 2.09 \times 10^{-9}$) (Fig. 4C). The sensitivity and specificity of the 13-TF signature was determined using time-dependent receiver operating characteristic (ROC) analysis, and the area under curve (AUC) at all follow-up time points were greater than 0.9 (Fig. 4D). The predictive model was then validated in a GEO dataset, and as shown in Fig. 5A, the high-risk group had worse survival compared to the low-risk group. In addition, the AUC values of the signature was greater than 0.75 (Fig. 5B). Taken together, the 13-TF signature can predict prognosis of ACC patients with high sensitivity and specificity.

### 3.3 The risk score is an independent prognostic factor of ACC

The prognostic relevance of the 13-TF signature was further validated by multivariate Cox regression analysis after normalizing for variables including age, gender and pathological stage. In both the training and validation ACC cohorts, the 13-TF risk score was an independent prognostic factor (Table 2). However, no significant correlation was seen between OS and age, gender or pathological stage. The high- and low-risk groups of both training and validation datasets were further divided into subgroups based on age ($\leq 50 \text{ vs } > 50$ years), gender (male vs female) and pathological stage ($\leq$ vs $>$. As shown in Fig. 6A and 6B, patients in the high-risk group had poor prognosis and significantly shorter survival compared to those in the low-risk group regardless of other variables. Thus, the 13-TF signature is an independent prognostic predictor of ACC.
### Table 2
Cox regression analysis of the correlation between clinicopathological factors and overall survival of ACC patients

<table>
<thead>
<tr>
<th>Variables</th>
<th>Group</th>
<th>Patients (N)</th>
<th>Univariate analysis</th>
<th>Patients (N)</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HR (95% CI)</td>
<td></td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td>TCGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk score</td>
<td>Low/High</td>
<td>40/39</td>
<td>16.95(5.02–57.2)</td>
<td>39/38</td>
<td>22.59(4.55–112.22)</td>
</tr>
<tr>
<td>Age</td>
<td>&lt;=50/&gt;50</td>
<td>41/38</td>
<td>1.80(0.85–3.82)</td>
<td>40/37</td>
<td>1.47(0.63–3.45)</td>
</tr>
<tr>
<td>Gender</td>
<td>F/M</td>
<td>48/31</td>
<td>1.00(0.47–2.14)</td>
<td>48/29</td>
<td>1.96(0.79–4.84)</td>
</tr>
<tr>
<td>Stage</td>
<td>I-II/III-IV</td>
<td>46/31</td>
<td>6.48(2.71–15.5)</td>
<td>46/31</td>
<td>1.78(0.72–4.39)</td>
</tr>
<tr>
<td>GSE19776</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk score</td>
<td>Low/High</td>
<td>11/11</td>
<td>16.57(2.02–135.76)</td>
<td>11/11</td>
<td>18.85(2.09–170.08)</td>
</tr>
<tr>
<td>Age</td>
<td>&lt;=50/&gt;50</td>
<td>15/7</td>
<td>1.72(0.60–4.88)</td>
<td>15/7</td>
<td>1.23(0.38–3.97)</td>
</tr>
<tr>
<td>Gender</td>
<td>F/M</td>
<td>11/11</td>
<td>1.26(0.46–3.44)</td>
<td>11/11</td>
<td>0.77(0.26–2.26)</td>
</tr>
<tr>
<td>Stage</td>
<td>I-II/III-IV</td>
<td>14/8</td>
<td>1.21(0.44–3.29)</td>
<td>14/8</td>
<td>1.46(0.51–4.14)</td>
</tr>
</tbody>
</table>

#### 3.4 The 13-TF signature is associated with cancer-related functions

GSEA results showed that four hallmarks including G2M_CHECKPOINT (P = 0.021), E2F_TARGETS (P = 0.023), SPERMATOGENESIS (P = 0.046) and MITOTIC_SPINDLE (P = 0.048) were significantly enriched in high-risk patients, suggesting a mechanistic basis of the prognostic role of 13-TF signature in ACC (Fig. 7).

#### 4 Discussion

ACC is a rare endocrine cancer with limited therapeutic options and poor clinical outcomes. Studies have increasingly identified molecular diagnostic and prognostic signatures of various cancers by screening multiple databases via high-throughput technologies such as microarrays and next-generation sequencing. Transcription factors (TFs) are often aberrantly expressed in tumors, and correlated with cancer prognosis. Recent studies have identified specific TFs that are independent prognostic factors in various cancers. The zinc-finger transcription factor Snail is associated with decreased survival of ACC patients and higher risk of distant metastasis. However, a TF-related prognostic signature has not yet been identified for ACC.

We analyzed the gene expression profiles of ACC patients deposited in GEO and TCGA databases, and constructed a prognostic signature of 13 TFs, including CREB3L3, NR0B1, CENPA, FOXM1, E2F2, MYBL2, HOXC11,
ZIC2, ZNF282, DNMT1, TCF3, ELK4 and KLF6. NR0B1, also known as DAX1, is an atypical orphan nuclear receptor that plays a key role in several cancers. NR0B1 silencing decreased the in vitro invasiveness of the lung cancer cell line A549 and inhibited xenograft growth without affecting cell proliferation. In addition, NR0B1 is overexpressed in cervical cancer and promotes cancer cell proliferation via the Wnt/β-catenin pathway. It is also overexpressed in adrenocortical tumors, ovarian cancer, breast cancer, endometrial cancer and prostate cancer, although its potential role in ACC has not been investigated.

CEPNA is a histone-H3 variant that regulates cell division by establishing kinetochore assembly and ensuring proper centromere segregation, and is associated with cancer progression. CENPA expression level is a potential biomarker of poor prognosis in cancer patients. FOXM1 and E2F2 are the upstream regulators of CENPA and play critical roles in cell cycle progression and tumorigenesis. Both TFs can potentially bind to the CENPA promoter sequence indicating that they regulate CENPA transcription. Previous studies have correlated CENPA with poor prognosis in ACC patients and identified E2F2 as an ACC-related TF. Although few studies have associated CREB3L3, MYBL2, HOXC11, ZIC2, ZNF282, DNMT1, TCF2, FLK4 and KLF6 with ACC progression, there is no evidence linking FOXM1 to ACC. We found that high levels of CREB3L3 and NR0B1 were correlated with good prognosis, while that of other TFs were correlated with poor prognosis in the ACC patients. Tumor stage was correlated with OS of patients in the training set but not in the validation set. Furthermore, the 13-TF signature was an independent prognostic factor in both TCGA and GEO datasets.

The GSEA results showed that G2M-CHECK POINT and E2F-TARGET were significantly enriched in the high-risk group. The G2/M checkpoint is frequently impaired in cancer cells, which promotes genomic instability and tumorigenesis. Since the E2F transcription factors regulate DNA replication and are aberrantly expressed in almost all human cancers, targeting E2Fs could be a generic approach in anti-cancer treatment.

Conclusions

To the best of our knowledge, this report is the first to investigate the cancer-specific TFs and their association with clinical outcomes in ACC patients. The 13-TF signature showed accurate predictive ability and is a promising prognostic biomarker for clinical applications. However, the in-silico results were not validated by PCR or Western blotting. Future studies should focus on validating these survival-related TFs through molecular and functional assays, and determine the mechanistic basis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data are available in the manuscript and they are exhibited in figures.
**Competing interests**

The authors declare that they have no competing interests.

**Funding**

Not applicable.

**Authors’ contributions**

JYZ and XPL: study design, drafting and revision the manuscript; JYZ and BL: acquisition and analysis of data. All the authors read and approved the final manuscript.

**Acknowledgements**

Not applicable.

**References**


Figures

Flowchart describing the construction and validation of the 13-TF prognostic signature.
Figure 2

Construction of the 13-TF prognostic signature. A, Selection of optimal tuning parameter (λ) in the LASSO model. The dotted vertical lines are drawn at the optimal values by minimum criteria (left) and 1-SE criteria (right). Error bars represent standard error (SE); B, Coefficient of each transcription factor; C, Spearman correlation analysis for the 13 TFs. Red and green respectively represent positive and negative correlation. The size of the point represents the strength of correlation.
Figure 3

Survival curves of ACC patients demarcated on the basis of TF risk score. The median risk score was used as the cutoff value. ACC, Adrenocortical carcinoma; OS, overall survival.
Figure 4

Risk score calculated by the 13-TF signature and Kaplan-Meier survival in TCGA dataset. A, Risk score distribution of the 13-TF signature and patient survival status; B, Heatmap of each TF; C, Kaplan-Meier survival plots; D, Prognostic accuracy of the 13-TF signature. HR, Hazard ratio; AUC, area under the curve.
Figure 5

Kaplan-Meier survival analysis of ACC patients in GEO (Gene Expression Omnibus) validation set. A, Kaplan-Meier plots for OS. B, ROC curve for predicting OS at different time points. HR, hazard ratio; ROC, receiver operating characteristic; AUC, area under the curve.
Figure 6

Clinical characteristics of TCGA and GEO cohorts stratified by the 13-TF risk score. A, TCGA set; B, GEO set. TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus.
Figure 7

Gene set enrichment analysis of 13 survival-related TFs.