EBioMedicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

A Genomic-clinicopathologic Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer

Shao-Xu Wu [a,e,1], Jian Huang [a,1], Zhuo-Wei Liu [b,1], Hai-Ge Chen [d,1], Pi Guo [f,1], Qing-Qing Cai [c,1], Jun-Jiong Zheng [a], Hai-De Qin [e], Zao-Song Zheng [a], Xin Chen [b], Rui-Yun Zhang [d], Si-Liang Chen [b], Tian-Xin Lin [a,*]

[a] Department of Urology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China
[b] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Department of Urology, Sun Yat-sen University Cancer Center, Guangzhou, China
[c] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Department of Medical Oncology, Sun Yat-sen University Cancer Center, Guangzhou, China
[d] Department of Urology, Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China
[e] Guangdong Province Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Research Center of Medicine, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China
[f] Department of Preventive Medicine, Shantou University Medical College, Shantou, China

ARTICLE INFO

Article history:
Received 19 March 2018
Received in revised form 29 March 2018
Accepted 29 March 2018
Available online xxxx

Keywords:
Genomics
Lymphatic metastasis
Nomogram
Urinary bladder neoplasms

ABSTRACT

Preoperative lymph node (LN) status is important for the treatment of bladder cancer (BCa). Here, we report a genomic-clinicopathologic nomogram for preoperatively predicting LN metastasis in BCa. In the discovery stage, 325 BCa patients from TCGA were involved and LN-status-related mRNAs were selected. In the training stage, multivariate logistic regression analysis was used to developed a genomic-clinicopathologic nomogram for preoperative LN metastasis prediction in the training set (SYSMH set, $n = 178$). In the validation stage, we validated the nomogram using two independent sample sets (SYSUCC set, $n = 142$; RJH set, $n = 104$) with respect to its discrimination, calibration and clinical usefulness. As results, we identified five LN-status-related mRNAs, including *ADRA1D*, *COL10A1*, *DKK2*, *HIST2H3D* and *MMP11*. Then, a genomic classifier was developed to classify patients into high- and low-risk groups in the training set. Furthermore, a nomogram incorporating the five-mRNA-based classifier, image-based LN status, transurethral resection (TUR) T stage, and TUR lymphovascular invasion (LVI) was constructed in the training set, which performed well in the training and validation sets. Decision curve analysis demonstrated the clinical value of our nomogram. Thus, our genomic-clinicopathologic nomogram shows favorable discriminatory ability and may aid in clinical decision-making, especially for cN-patients.

1. Introduction

Bladder cancer (BCa) is a common cancer with globally high mortality (Burger et al., 2013; Antoni et al., 2017). Lymph nodes (LNs) are the most common metastatic sites in BCa. Previous studies suggested that 25%–30% of BCa patients treated with radical cystectomy (RC) and pelvic lymph node dissection (PLND) underwent LN metastasis (Stein et al., 2001; Leissner et al., 2004; Vazina et al., 2004; Abol-Enein et al., 2011; Baltaci et al., 2011; Zehnder et al., 2011; Jensen et al., 2012). Up to 80% of BCa patients with pathologic LN metastasis suffer from recurrence after undergoing RC, while only approximately 30% of BCa patients who are LN-negative (pN0) experience tumor recurrence (Shariat et al., 2006; Stamatakis et al., 2012). In addition, LN-positive (pN1–3) patients have a significantly lower five-year overall survival rate compared with pN0 patients (15%–31% vs. > 60%) (Bassi et al., 1999; Stein et al., 2001; Karl et al., 2009; Zehnder et al., 2014).

Preoperative LN status is critical for BCa treatment decision-making, particularly in helping determine the extent of PLND and the use of neoadjuvant chemotherapy (Kluth et al., 2015; Zargar-Shoshtari et al., 2016). Currently, contrast-enhanced computed tomography (CT) is the standard clinical procedure for preoperatively evaluating the LN stage (McKibben and Woods, 2015). However, the sensitivity of CT at detecting metastatic lesions in the LNs is relative low (31%–45%) (Baltaci et al., 2008; Lodde et al., 2010; Goodfellow et al., 2014). As a consequence, a considerable portion of patients with inaccurate staging may receive inadequate treatment or overtreatment. In particular, understaging might lead to postoperative recurrence or even death (Culp et al., 2014); conversely, overstaging is likely to subject a patient to needless neoadjuvant chemotherapy or unnecessarily extensive PLND. Therefore, there is an urgent need to improve the nodal staging accuracy for the treatment of BCa.

* Corresponding author.
E-mail address: lintx@mail.sysu.edu.cn. (T.-X. Lin).
[1] These authors contributed equally

https://doi.org/10.1016/j.ebiom.2018.03.034

In recent years, gene expression signatures have been used to predict LN metastasis in BCa (Smith et al., 2011; Seiler et al., 2016). However, the models developed in these studies have not yet been used in clinical practice due to limitations. Smith et al. developed a 20-mRNA-based classifier for predicting LN status preoperatively in BCa patients (Smith et al., 2011). The predictive efficacy of the model was modest, with an area under the curve (AUC) of 0.67 in the external validation set. Seiler et al. developed a 51-RNA-based classifier that achieved an AUC of 0.82 for predicting BCa LN metastasis (Seiler et al., 2016), but this study was limited by the lack of external validation. Moreover, the previously developed models were limited to clinically node-negative (cN-) BCa patients, and clinical factors were not incorporated into the predictive models for evaluation.

In the current study, we postulated that inclusive models incorporating a genomic signature and clinicopathologic factors might improve the accuracy of nodal staging. We identified mRNAs that significantly correlated with LN metastasis by mining RNA-SEQ data from the TCGA BLCA project and then developed a multiple-mRNA classifier in our BCa sample set. The genomic classifier was further combined with clinicopathological factors to build an inclusive nomogram for predicting LN status preoperatively. We assessed the predictive accuracy of the nomogram and validated it in two independent sample sets. We also evaluated the predictive efficacy of the nomogram in clinically low-risk subgroups (non-muscle invasive bladder cancer [NMIBC] or cN-).

2. Materials and Methods

2.1. Patients and Clinical Database

In this study, a total of 424 BCa patients who had undergone RC and PLND without preoperative therapy were recruited from three independent cancer centers, including 178 samples from the Sun Yat-sen Memorial Hospital of Sun Yat-sen University (SYSMH) between March 2006 and December 2017, 142 samples from the Sun Yat-sen University Cancer Center (SYSUCC) between April 2002 and September 2017, and 104 samples from the Renji Hospital of Shanghai Jiaotong University School of Medicine (RJH) between June 2013 and July 2017. The inclusion criteria were pT0-4N0-3M0 BCa patients who underwent RC + PLND and those with tumor samples, which were confirmed urothelial carcinoma pathologically. The exclusion criteria were BCa patients who underwent preoperative therapy (either neoadjuvant chemotherapy or radiotherapy) or didn't receive pre-RC transurethral resection of bladder tumor (TURBT) in the three centers, and those with samples with insufficient total RNA or a failed quality control (QC) step. All patients underwent CT or magnetic resonance imaging (MRI) before TURBT. The procedures of PLND and RNA QC were described in the Supplementary Material in detail. The patient recruitment pathway is shown in Supplementary Fig. S1. Clinicopathological data were collected through medical record review and included age, gender, BCa recurrence, image-based tumor size, image-based tumor number, image-based N stage, status of hydronephrosis, TUR T stage, TUR tumor grade, and TUR lymphovascular invasion (LVI). LVI was considered present only if tumor cells were unequivocally presented within or attached to the wall of a vascular or lymphatic space on hematoxylin & eosin stained sections (Cho et al., 2009). As for indeterminate cases and aggressive tumors cases, multiple serial sections were used. The BCa patients were classified using the 2009 TNM staging system (Sobin et al., 2009) and the 2004 WHO classification (Epstein et al., 2004). We defined cases as clinical LN positive (cN+) if pelvic LN > 8 mm or abdominal LN > 10 mm in the maximum short-axis diameter based on CT or MRI (Dorfman et al., 1991; Barentsz et al., 1999). To ensure validity of the pathologic outcomes extraction, all samples were reassessed by two pathologists (Hong Zen and Lin Wang) while blinded to patient clinicopathological data and the findings of the other reviewer. Interreader reliability measured using the intraclass correlation coefficient was >0.95 for each pathologic characteristic. We used the SYSMH samples as the training set for model development and the SYSUCC and RJH samples as the external validation sets. The institutional review boards of the three centers approved this study and the need to obtain informed consent was waived.

2.2. qRT-PCR

Total RNA from 424 fresh-frozen BCa tissue samples was extracted using RNAiso plus reagent (TaKaRa) according to the manufacturer's instructions, and the expression of BCa LN status-related mRNAs (selected in the discovery stage) was further examined via qRT-PCR. First-strand cDNA was synthesized with PrimeScript™ RT Master Mix (TAKARA) according to the manufacturer's instructions. The qRT-PCR was conducted to examine the expression of the selected mRNAs using SYBR-Green PCR Master Mix (Roche) on a LightCycler 96 Real-Time PCR instrument (Roche). *Homo sapiens* actin beta (ACTB) was used as an internal reference gene to normalize mRNA levels between different samples for an exact comparison of transcript level. Expression levels of each mRNA were calculated using the $-\triangle CT$ approach ($\triangle CT$ = CT of mRNA – CT of ACTB RNA).

2.3. Source of the TCGA BLCA Project Data

Normalized gene expression data (level 3, RNA-SEQ data from an Illumina HiSeq 2000 platform) were obtained from the UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu) on 1 June 2016. The UCSC website contains a detailed description of data normalization. Briefly, the gene expression profile was measured experimentally using the Illumina HiSeq 2000 RNA Sequencing platform at a TCGA Genome Center. Level 3 interpreted data were downloaded from the TCGA Data Portal (https://portal.gdc.cancer.gov/). This dataset provides gene-level transcription estimates as an RSEM normalized count. Genes were mapped onto the human genome coordinates using the UCSC cgData HUGO probeMap.

3. Procedures

Our study was conducted in three stages: discovery stage, training stage and validation stage. The study flowchart is presented in Fig. 1. In the discovery stage, we used paired samples (tumor/normal) from BCa patients in the TCGA BLCA project as discovery set I ($n = 19$ pairs); all BCa patients with nodal status information in the TCGA were included in discovery set II ($n = 325$). Discovery set II included 113 patients with LN-positive disease (pN1–3) and 212 patients with LN-negative disease (pN0). A volcano plot analysis of discovery set I was used to screen for differentially expressed mRNAs, and we then used the least absolute shrinkage and selection operator (LASSO) logistic regression algorithm (Tibshirani, 1996) to identify BCa-associated mRNAs as candidates. Furthermore, the LASSO logistic regression model was used to screen LN status-correlated mRNAs from the candidate mRNAs in discovery set II (see Fig. 2 and Statistical Analysis section).

In the training stage, a multivariable logistic regression model was used to construct a multi-mRNA-based classifier for predicting LN status in the training (SYSMH) set based on the significant LN status-correlated mRNAs identified in the discovery stage. We then derived the genetic risk scores from the multi-mRNA-based classifier. Receiver operating characteristic (ROC) analysis was performed to investigate the predictive efficiency of the multi-mRNA-based classifier by measuring the AUC. Furthermore, independent validations were conducted using two external samples sets, the SYSUCC and RJH sets. In addition, we evaluated the classification performance of the model in distinguishing BCa patients with a low or high risk of LN metastasis.

To calculate the genetic risk scores reflecting the risk of LN metastasis for each patient, a regression equation was derived using the estimated coefficients in the above multivariable regression model (the

Please cite this article as: Wu, S.-X., et al., A Genomic-clinicopathologic Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer, EBioMedicine (2018), https://doi.org/10.1016/j.ebiom.2018.03.034