

Research Article

An Epithelial-Mesenchymal Transition (EMT) Preoperative Nomogram for Prediction of Lymph Node Metastasis in Bladder Cancer (BLCA)

Rui Cao,¹ Bo Ma,² Gang Wang,³ Yaoyi Xiong,⁴ Ye Tian,¹ and Lushun Yuan ⁵

¹Department of Urology, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China

²Department of Stomatology, Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China

³Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

⁴Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

⁵Department of Internal Medicine, Division of Nephrology, Leiden University Medical Center, Leiden, 2333 ZA, Netherlands

Correspondence should be addressed to Lushun Yuan; l.yuan@lumc.nl

Received 19 July 2020; Revised 28 September 2020; Accepted 14 October 2020; Published 3 November 2020

Academic Editor: Jian Ma

Copyright © 2020 Rui Cao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lymph node (LN) metastasis is a lethal independent risk factor for patients with bladder cancer (BLCA). Accurate evaluation of LN metastasis is of vital importance for disease staging, treatment selection, and prognosis prediction. Several histopathologic parameters are available to predict LN metastasis postoperatively. To date, medical imaging techniques have made a great contribution to preoperatively diagnosis of LN metastasis, but it also exhibits substantial false positives. Therefore, a reliable and robust method to preoperatively predict LN metastasis is urgently needed. Here, we selected 19 candidate genes related to epithelial-mesenchymal transition (EMT) across the LN metastasis samples, which was previously reported to be responsible for the subtype transition and correlation with malignancy and prognosis of BLCA, to establish an EMT-LN signature through LASSO logistic regression analysis. The EMT-LN signature could significantly predict LN metastasis with high accuracy in the TCGA-BLCA cohort, as well as several independent cohorts. As integrating with C3orf70 mutation, we developed an individualized prediction nomogram based on the EMT-LN signature. The nomogram exhibited good discrimination on LN metastasis status, with AUC of 71.7% and 75.9% in training and testing datasets of the TCGA-BLCA cohort. Moreover, the EMT-LN nomogram displayed good calibration with $p > 0.05$ in the Hosmer-Lemeshow goodness of fit test. Decision curve analysis (DCA) revealed that the EMT-LN nomogram was of high potential for clinical utility. In summary, we established an EMT-LN nomogram integrating an EMT-LN signature and C3orf70 mutation status, which acted as an easy-to-use tool to facilitate preoperative prediction of LN metastasis in BLCA individuals.

1. Introduction

Bladder cancer (BLCA) is urogenital malignancy with high mortality and morbidity, which is usually divided into two major subtypes: non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) [1, 2]. About 70% of new BLCA diagnosis cases are NMIBC, which are not that life-threatening but have a potential to recur and progress even though receiving repeated transurethral resections of bladder tumours (TURBTs) and intravesical therapies with Bacillus Calmette-Guerin (BCG) or

chemotherapeutic drugs [3]. Moreover, patients with pathological T1G3 NMIBC or contaminated with carcinoma in situ (CIS) are more likely to develop into MIBC or exhibit lymph node metastasis [4]. Although only 30% of BLCA patients are diagnosed with MIBC at first diagnosis, they are responsible for the vast majority of BLCA-specific deaths. The standard treatment procedure for localized MIBC is radical cystectomy (RC) accompanied by cisplatin-based neoadjuvant chemotherapy (NAC), adjuvant chemotherapy (AC), or in selected cases trimodal therapy (TURBTs, chemotherapy, and radiotherapy) [5, 6]. Moreover, patients with

advanced or metastatic BLCA are recommended for cisplatin-based chemotherapy or the novel immune checkpoint inhibitor (ICI) immunotherapy, which were found to improve the prognosis of patients with advanced malignancy, such as non-small-cell lung cancer (NSCLC), melanoma, and BLCA [7–10]. But we have to notice that the response rate of immunotherapy is only modestly above the historical 10% response rate to traditional chemotherapies.

Lymph node (LN) metastasis is an independent risk factor for BLCA that patients with LN metastasis demonstrated a poorer prognosis compared with patients without lymphatic spread [11]. As pelvic is the most important route within the draining lymph nodes of BLCA, mounting evidence suggests that addition of pelvic lymph node dissection (PLND) and/or extended lymphadenectomy to RC showed substantial oncological benefit compared with non-PLND cohorts, irrespective of pathological nodal status [12, 13]. And PLND was highlighted by providing prognostic information such as tumour burden, lymph node density, and extracapsular extension of metastatic LNs, which can guide subsequent treatment management [14, 15]. Currently, various PLND templates are used in clinical practice while extended PLND templates were reported to provide optimal recurrence-free and cancer-specific survival when compared with standard PLND and superextended PLND [16, 17]. However, the potential drawbacks, such as increased operative time and adjacent visceral, vascular, and lymphocele injury, must be weighed against those potential benefits. Moreover, nearly 50% of patients with MIBC already exhibited occult distant metastases at the time of diagnosis [18]. Peyton et al. reported that NAC followed by RC could induce downstaging and significantly improve prognosis compared with RC alone for patients with BLCA [19]. But Cha et al. found that LN-positive patients previously treated with NAC exhibited a strikingly poorer prognosis than LN-positive patients subsequently treated with AC and should be considered for regimens using trimodal therapy approaches or novel agents [20]. Therefore, accurate evaluation of LN metastasis in BLCA patients is of vital importance for disease staging, treatment selection, and prognosis prediction.

As the bladder has a multitude of primary lymphatic landing sites, lymphoscintigraphy and the sentinel LN concept have limited value for the detection of regional LN metastases in BLCA, which has been confirmed by several clinical studies [21, 22]. Currently, many imaging techniques could help the surgeons to remove suspicious nodes which could not be resected in a standardized template. CT and conventional MRI, which mainly rely on morphologic criteria including LN size, shape, and internal architecture, are widely used for preoperative detection of LN metastases in patients with BLCA. Until now, there is a consequent lack of consensus regarding the normal limit for size in the diagnosis of pelvic LN metastases as the size of nonmetastatic LNs varies widely and may overlap with the size of LN metastases [23]. In BLCA, their diagnostic accuracy is less than optimal as CT might miss 30% to 40% of LN metastases, resulting in upstaging which pathological N+ is frequently found despite negative preoperative imaging [24, 25]. Furthermore, positron emission tomography (PET) depends on

the increased metabolic rate of specific tissues and their volume to improve its diagnosis efficiency of malignancy, which is different from conventional CT or MRI just relied on traditional parameters such as size or shape. But a meta-analysis demonstrated low sensitivity rates of choline PET/CT (pooled sensitivity and pooled specificity of 49% and 95%) in the setting of LN staging prior to RC [26]. The main cause of low sensitivity might be that only small amounts of the tracer are accumulated in small metastases, which are probably insufficient for positive imaging and are likely to be missed. Moreover, some groups even used molecular markers, such as MUC7 and uroplakin II, to detect occult LN metastases, but the accuracy efficiency is not that stable [27, 28].

With the rapid development of next-generation sequencing technology, we could comprehensively investigate the transcriptome and genetic alteration in individuals to identify the difference of BLCA in terms of LN metastasis status at a genomic level. The last decades have witnessed a transition away from multiple biomarkers from individual analyses to combined analysis of a panel of biomarkers for construction of a signature, which was recognized as a promising and powerful approach for clinical management [29]. Kessel et al. even established a gene signature to predict LN metastasis in MIBC before RC, but their model showed low AUC with no clinical applicability [30]. In our previous study, we found that epithelial-mesenchymal transition (EMT), which is a multistep process where epithelial cells lose their epithelial characteristics and gain mesenchymal characteristics, such as motility and invasive properties, is responsible for the subtype transition from NMIBC to MIBC and significantly correlated with BLCA malignancy and prognosis [31]. Therefore, in the present study, we aim to establish a nomogram incorporating an EMT-LN signature and somatic genetic mutations for preoperative prediction of LN metastasis in BLCA.

2. Materials and Methods

2.1. Data Collection and Processing. The TCGA-BLCA dataset, which was utilized as the training and internal validation cohort, was obtained from the TCGA Genomic Data Commons (GDC) data portal (<https://portal.gdc.cancer.gov/>). The Level 3 transcriptome HTSeq-counts and fragments per kilobase per million (FPKM) data of the TCGA-BLCA cohort were downloaded from the TCGA Genomic Data Commons (GDC) data portal (<https://portal.gdc.cancer.gov/>). Since multiple ENSEMBL IDs mapped to a single gene symbol, the highest expressed ENSEMBL ID was used. Transcripts per kilobase million (TPM) values were transformed from the FPKM values to represent the gene expression, which is more similar to gene expression from microarrays and more comparable between samples [32]. Detailed information of clinicopathological characteristics in the TCGA-BLCA cohort can be found in our previous study [31]. Then, we have enrolled 128 samples with LN metastasis (LN+) and 228 samples without LN metastasis (LN-) as the entire TCGA-BLCA cohort. Three independent microarray cohorts including GSE13507, GSE31684, and GSE106534 were obtained for external validation from Gene Expression

Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>). The raw data were processed via RMA algorithm background correction, log₂ transformation, quantile normalization, and annotation by the package “*Affy*” in R [33]. When several probes mapped to a single gene symbol, the highest expressed probe was annotated as the gene expression. Data were analysed with the R (version 3.5.3) and Bioconductor packages.

2.2. Differential Expression and Functional Enrichment Analyses. Differential expression analysis was performed by comparing LN+ and LN- samples in the TCGA-BLCA cohort through package “*DESeq2*” in R, which contains multiple testing with an embedded Benjamini-Hochberg procedure for *p* value adjustment. The significance criteria for determining DEGs were set as false discovery rate (FDR) < 0.05 and |log₂ 2FC| > 1.0. Then, gene set enrichment analysis (GSEA) was performed to investigate the difference on biological processes between LN+ and LN- samples through package “*clusterProfiler*” in R [34, 35]. We also utilized a single-cell gene set enrichment analysis (ssGSEA) algorithm to validate the difference of EMT pathway enrichment in individuals based on gene sets of “h.all.v7.1.symbols,” which was downloaded from MSigDB of Broad Institute (<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>). To further investigate the relationship between EMT pathways and LN status, more EMT-related pathways were used for GSEA analysis, and EMT-related gene sets were collected via key words “epithelial-mesenchymal transition” or “EMT” in the GSEA website (<https://www.gsea-msigdb.org/gsea/msigdb/search.jsp>).

2.3. Somatic Genetic Mutation Analysis. The TCGA-BLCA somatic mutation data processed with the MuTect2 algorithm were downloaded from the Genomic Data Commons (<https://portal.gdc.cancer.gov/>) using the package “*TCGAbiolinks*” in R [36]. The significant tumour mutated genes (driver genes) were identified through “*MutSigCV_v1.41*” (20) (<http://www.broadinstitute.org>), which was a reliable algorithm for mining the driver genes in disease development and has included the following features: high mutational burden relative to background expectation, accounting for heterogeneity, clustering of mutations within the gene, and enrichment of mutations in evolutionarily conserved sites. The significance levels (*p* values) from each test were combined to obtain a single significance level per gene [37]. The procedure was processed with default parameters, and *q* value was set as < 0.05. The oncoprint of mutation landscape was visualized by package “*ComplexHeatmap*” in R [38]. Then, all the significant frequent nonsilent mutation genes were validated by an independent test between the LN+ and LN- groups with *p* < 0.05.

2.4. Generation of EMT-LN Signature. Total 228 LN- and 128 LN+ samples in the TCGA-BLCA cohort were randomly separated into two datasets based on 10-fold stratified sampling, where the training dataset included 9-fold LN+ and LN- samples and the testing dataset included the other 1-fold sample. Then, the EMT-related genes were screened out through the DEGs and identified as the candidate

EMT-DEGs for further analysis. Least absolute shrinkage and selection operator (LASSO) logistic regression analysis based on the package “*glmnet*” in R was performed to select primary predicative features and build an EMT-LN signature [39]. The optimal values of the penalty parameter λ were determined through tenfold cross-validation error for dimension reduction to reduce noise or redundant genes. The risk score of the EMT-LN signature for each sample was calculated via a linear combination of the selected features and weighted by the corresponding coefficients. The formula of risk score = $\sum_{i=1}^n (\text{coef}_i \times \text{Expr}_i)$, where Expr_i is the relative abundance of the feature in the signature for patient *i* and coef_i is the corresponding coefficients of feature *i*. The correlation between EMT-LN signature and LN metastasis status was then evaluated in training the testing datasets by using the Mann-Whitney *U* test.

2.5. Development of an Individualized Prediction Model. The EMT-LN signature and significant tumour frequent mutations were selected as candidate features for univariable and multivariable logistic regression analysis. Then, features with respective *p* < 0.05 were retained in the prediction model. An EMT-LN signature-incorporated nomogram was generated with the packages “*rms*,” “*nomogramEx*,” and “*regplot*” in R as a quantitative tool for clinicians to predict the LN metastasis probability for individuals.

2.6. Validation of EMT-LN Signature and Nomogram. After construction of the EMT-LN signature, internal validation was performed in the testing dataset of the TCGA-BLCA cohort based on the formula of the risk score established in the training dataset. Then, the receiver operating characteristic (ROC) curve and calibration curves were used to detect the prediction accuracy and stability of the nomogram by using packages “*pROC*” and “*rms*” in R. Moreover, external validation for the EMT-LN signature was also tested in the three independent GEO cohorts (GSE13507, GSE31684, and GSE106534). Since the mutation data were absent in these cohorts, unsupervised clustering was harnessed to determine if the EMT-LN signature could help us to distinguish LN metastasis status and whether it was associated with prognosis in three GEO cohorts. A supervised hierarchical clustering method (Ward.D) for analysis of the EMT-LN signature was applied to identify clusters with *k* = 2 based on 1-Pearson’s correlation distance. Before supervised hierarchical clustering, expression profiling was transformed by log₂(*x* + 1) and median-centered.

2.7. Statistical Analyses. The statistical significance of continuous variables between two groups was estimated by Student’s *t*-test or Mann-Whitney *U* test when appropriate. For categorical variables, Fisher’s exact or χ^2 tests were used. Kaplan-Meier curves and Cox regression for survival analysis were generated which is performed by package “*survival*” and “*survminer*” in R [40]. The significant differences between survival curves belonging to different defined groups were determined with the log-rank (Mantel-Cox) test. Multivariate Cox proportional hazard models were used to estimate the hazard ratios of variables and determine independent

prognostic factors. The oncoprint was used to present the mutation landscape of significantly mutated genes in the TCGA-BLCA cohort via packages “*mafTools*” [41] and “*complexheatmap*” [38] in R. A nomogram and calibration curves were built with the packages “*rms*,” “*nomogramEx*,” and “*regplot*” in R. DCA was performed to determine whether our established nomogram was of clinical usefulness according to Iasonos et al.’s suggestion [42]. The package “*pROC*” in R was used to plot and visualize receiver operating characteristic (ROC) curves. All statistical analyses were performed with R software 3.5.3. Statistical significance was set at $p < 0.05$.

3. Results

3.1. Demographic Characteristics. The design and workflow of the study are shown in Figure S1. We comprehensively evaluated the difference in clinicopathological characteristics based on LN metastasis status. The distributions of gender and age (dichotomized by median age of 65) were not different between LN+ and LN- samples. But we observed that the histology subtype ($p = 0.0237$), grade ($p = 2.92E - 22$), lymph nodes positive by hematoxylin and eosin (HE) (dichotomized by the number of 0, $p = 6.41E - 36$), lymphovascular invasion ($p = 5.40E - 14$), AJCC pathological T stage ($p = 1.07E - 06$), AJCC pathological N stage ($p = 8.35E - 50$), AJCC pathological M stage ($p = 8.91E - 17$), AJCC pathological tumour stage ($p = 9.765503E - 71$), extracapsular extension ($p = 0.002231268$), and tumour status ($p = 1.05E - 13$) were significantly associated with LN metastasis status (Table 1). Moreover, the LN metastasis status was also significantly correlated with survival-related parameters, such as the primary therapy outcome ($p = 1.13E - 05$), additional treatment outcome ($p = 2.59E - 06$), and vital status ($p = 2.02E - 07$) (Table 1). As expected, Kaplan-Meier survival curves demonstrated that patients with LN+ exhibited poorer overall survival (OS) than LN- patients ($p < 0.001$, HR = 2.23, 95%CI = [1.6-3.12], Figure 1(a)), and a similar tendency could also be found in the recurrence-free survival (RFS) estimation where LN+ patients presented a higher recurrence rate ($p < 0.001$, HR = 2.59, 95%CI = [1.55-4.33], Figure 1(b)).

3.2. Differential Expression Genes (DEGs) between LN+ and LN- Tumours Are Enriched in EMT Signalling Pathway. The cut-offs were set as follows: $|\log_2(\text{fold change})| > 1$, false discovery rate (FDR) < 0.05 , and 352 differentially expressed genes (DEGs) were screened out between LN+ and LN- tumours by using the package “*DESeq2*” in R (Figure 2(a); Supplementary Table S1). Then, GSEA showed a striking higher enrichment of “Hallmark epithelial mesenchymal transition” gene set in LN+ tumours than that in LN- tumours (Figure 2(b); Supplementary Table S2). Moreover, we enrolled other gene sets associated with the EMT signalling pathway including “Jechlinger epithelial mesenchymal transition up” and “Gotzmann epithelial mesenchymal transition up” and found that they were also highly enriched in LN+ tumours

compared with LN- tumours (Figure 2(c); Supplementary Table S3). The patients with LN+ demonstrated a higher ssGSEA score of “Hallmark epithelial mesenchymal transition” gene set than LN- patients (Figure 2(d)).

3.3. Somatic Mutation Landscape between LN+ and LN- Tumours. After filtering out nonsilent mutation, 27 significantly mutated genes (SMGs, $q < 0.05$) were identified through all the 366 samples by utilizing the “*MutSigCV*” algorithm (Supplementary Table S4). Then, oncoprint visualized that most of SMGs were the driver genes in BLCA development, which has been reported before (Figure 2(e)) [43]. Moreover, we found that TP53, FGFR3, and C3orf70 within the SMGs were remarkably highly mutated in LN+ tumours when compared with LN- tumours through Fisher exact tests (Figures 2(f) and 2(g); Supplementary Table S5). Therefore, mutation of TP53, FGFR3, and C3orf70 was selected for further analysis.

3.4. Establishment of EMT-LN Signature. As the EMT signalling pathway was highly enriched in LN+ tumours, genes within the “Hallmark epithelial mesenchymal transition” gene set and the DEGs were merged as EMT-related candidate genes. Then, we submitted them to the LASSO logistic regression analysis and construct a signature (EMT-LN signature) for distinguishing LN metastasis status in the training dataset of the TCGA-BLCA cohort (Figures 3(a) and 3(b)). The coefficient of each feature and formula of the EMT-LN signature are shown in Supplementary Table S6. Moreover, the EMT-LN signature was strikingly higher in LN+ tumours than that in LN- tumours both in the training ($p = 6.2E - 10$) and testing (internal validation) datasets ($p = 0.012$) of the TCGA-BLCA cohort (Figures 3(c) and 3(d)).

3.5. Supervised Hierarchical Clustering by Using EMT-LN Signature. Supervised hierarchical clustering was performed to validate whether the EMT-LN signature could distinguish LN metastasis status in independent cohorts. In the GSE106534 cohort, we found that cluster C1 was almost accumulated in LN- tumours, while cluster C2 concentrated in LN+ patients (Figure 4(a)). Moreover, we observed that the EMT-LN signature was significantly correlated with recurrence-free survival (RFS), progression-free survival (PFS), and cancer-specific survival (CSS) where patients with cluster C1 were more likely to recur, progress, and decease in the GSE13507 cohort (Figure 4(b)). In the GSE31684 cohort, the similar role of the EMT-LN signature in disease metastasis survival (DMS), lymph node recurrence survival (LRS), and urinary tract recurrence survival (URS) was observed (Figure 4(c)). Furthermore, Kaplan-Meier survival curves revealed that the EMT-LN signature was associated with OS ($p = 0.129$) and CSS ($p = 0.004$) in the GSE13507 cohort (Figures 4(d) and 4(e)), as well as OS ($p = 0.073$) and RFS ($p = 0.099$) in the GSE3684 cohort (Figures 4(f) and 4(g)).

3.6. Development and Validation of an Individualized Prognostic Nomogram. Logistic regression analysis was performed to identify the preoperative features in predicting the LN metastasis status. As the pathological stage and

TABLE 1: Summary of detailed clinical information of the TCGA-BLCA cohort.

TCGA-BLCA	LN- (<i>n</i> = 228)	LN+ (<i>n</i> = 128)	Total (<i>n</i> = 356)	<i>p</i> value
Gender				0.8016887
Female	61 (26.8%)	32 (25.0%)	93 (26.1%)	
Male	167 (73.2%)	96 (75.0%)	263 (73.9%)	
Age				0.9572861
≤65	93 (40.8%)	50 (39.1%)	143 (40.2%)	
>65	135 (59.2%)	78 (60.9%)	213 (59.8%)	
Vital status				2.02E – 07*
Alive	153 (67.1%)	47 (36.7%)	200 (56.2%)	
Dead	75 (32.9%)	81 (63.3%)	156 (43.8%)	
Primary therapy outcome				1.13E – 05*
CR	157 (68.9%)	43 (33.6%)	200 (56.2%)	
PR	9 (3.9%)	10 (7.8%)	19 (5.3%)	
SD	12 (5.3%)	11 (8.6%)	23 (6.5%)	
PD	25 (11.0%)	33 (25.8%)	58 (16.3%)	
Additional treatment outcome				2.59E – 06*
CR	132 (57.9%)	28 (21.9%)	160 (44.9%)	
PR	2 (0.9%)	3 (2.3%)	5 (1.4%)	
SD	5 (2.2%)	6 (4.7%)	11 (3.1%)	
PD	42 (18.4%)	43 (33.6%)	85 (23.9%)	
Subtype				0.0237*
Nonpapillary	142 (62.3%)	95 (74.2%)	237 (66.6%)	
Papillary	83 (36.4%)	31 (24.2%)	114 (32.0%)	
Lymph nodes positive by HE				6.41E – 36*
>0	166 (72.8%)	2 (1.6%)	168 (47.2%)	
0	0 (%)	115 (89.8%)	115 (32.3%)	
Lymphovascular invasion				5.40E – 14*
No	96 (42.1%)	16 (12.5%)	112 (31.5%)	
Yes	53 (23.2%)	85 (66.4%)	138 (38.8%)	
AJCC pathological T stage				1.07E – 06*
T2	87 (38.2%)	17 (13.3%)	104 (29.2%)	
T3	105 (46.1%)	75 (58.6%)	180 (50.6%)	
T4	22 (9.6%)	33 (25.8%)	55 (15.4%)	
AJCC pathological N stage				8.35E – 50*
N0	228 (100.0%)	1 (0.8%)	229 (64.3%)	
N1	0 (%)	45 (35.2%)	45 (12.6%)	
N2	0 (%)	75 (58.6%)	75 (21.1%)	
N3	0 (%)	7 (5.5%)	7 (2.0%)	
AJCC pathological M stage				8.91E – 17*
M0	129 (56.6%)	45 (35.2%)	174 (48.9%)	
M1	0 (%)	7 (5.5%)	7 (2.0%)	
AJCC pathologic tumour stage				9.765503E – 71*
II	99 (43.4%)	1 (0.8%)	100 (28.1%)	
III	125 (54.8%)	1 (0.8%)	126 (35.4%)	
IV	3 (1.3%)	125 (97.7%)	128 (36.0%)	
Grade				2.92E – 22*
High grade	209 (91.7%)	126 (98.4%)	335 (94.1%)	
Low grade	18 (7.9%)	0 (%)	18 (5.1%)	

TABLE 1: Continued.

TCGA-BLCA	LN- (<i>n</i> = 228)	LN+ (<i>n</i> = 128)	Total (<i>n</i> = 356)	<i>p</i> value
Tumour status				1.05E – 13*
Tumour-free	152 (66.7%)	30 (23.4%)	182 (51.1%)	
With tumour	58 (25.4%)	79 (61.7%)	137 (38.5%)	
Extracapsular extension				0.002231268*
No	48 (21.1%)	35 (27.3%)	83 (23.3%)	
Yes	20 (8.8%)	48 (37.5%)	68 (19.1%)	
Mutation in TP53				0.027352667*
No	133 (57.5%)	59 (44.5%)	188 (52.8%)	
Yes	95 (42.5%)	69 (55.5%)	168 (47.2%)	
Mutation in FGFR3				0.037310566*
No	190 (83.8%)	117 (89.8%)	306 (86.0%)	
Yes	38 (16.2%)	11 (10.2%)	50 (14.0%)	
Mutation in C3orf70				0.013517206*
No	214 (93.9%)	27 (96.4%)	241 (94.1%)	
Yes	14 (6.1%)	1 (3.6%)	15 (5.9%)	

*Fisher's exact test $p < 0.05$.

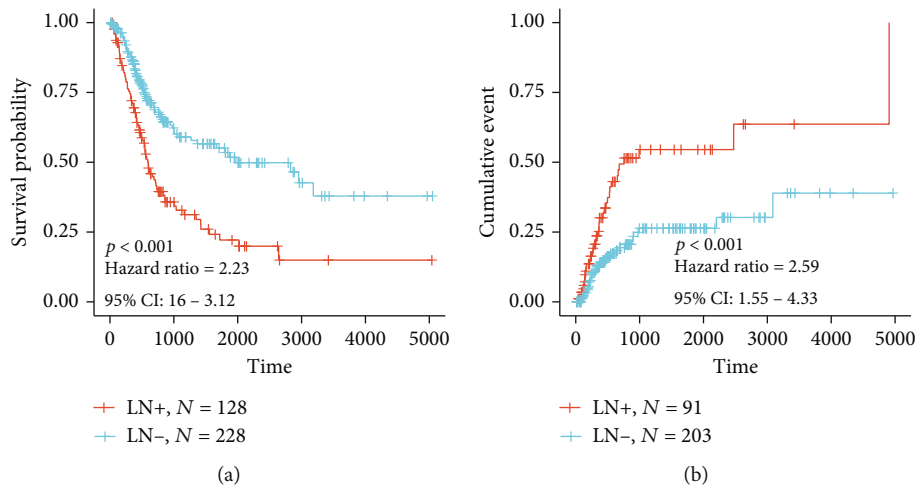


FIGURE 1: Association between LN metastasis status and prognosis of BLCA patients in the TCGA-BLCA cohort for overall survival (OS) (a) and for recurrence-free survival (RFS) (b). The difference in the prognosis of LN metastasis status was measured and visualized through Kaplan-Meier survival curves and determined with the log-rank (Mantel-Cox) test.

detailed TNM classification should mostly be detected by biopsy postoperatively not preoperatively, we just enrolled the EMT-LN signature and mutation of TP53, FGFR3, and C3orf70 as the candidate features to establish the preoperative nomogram. We found that only the EMT-LN signature and C3orf70 mutation have statistical significance in the full model with $p < 0.05$ in the logistic regression (Supplementary Table S7). Hence, we integrated these two features to develop an EMT-LN nomogram (Figure 5). According to the nomogram, every patient would obtain a total point value by adding the point for each prognostic parameter. And higher total points correspond to a higher presence of LN metastasis. Then, the calibration curve of the nomogram showed that the performance of our model was similar to

the ideal model in the training dataset, with a Hosmer-Lemeshow test suggesting no departure from perfect fit ($p > 0.05$). ROC curves revealed that the EMT-LN nomogram was efficient in predicting the LN metastasis status, with high AUCs in both training dataset (AUC: 0.717 (0.659–0.776)) and testing dataset (AUC: 0.759 (0.603–0.915)) (Figure 6(b)). Moreover, external validation also obtained AUCs of 0.64, 0.612, and 0.619 in GSE10634, GSE32684, and GSE13507, respectively, when deploying the EMT-LN nomogram (Figure S2). Then, decision curve analysis (DCA) demonstrated that the EMT-LN nomogram exhibited a high net benefit over the “treat-all” or “treat-none” strategies, which indicated that the EMT-LN nomogram had high potential clinical utility (Figure 6(c)).

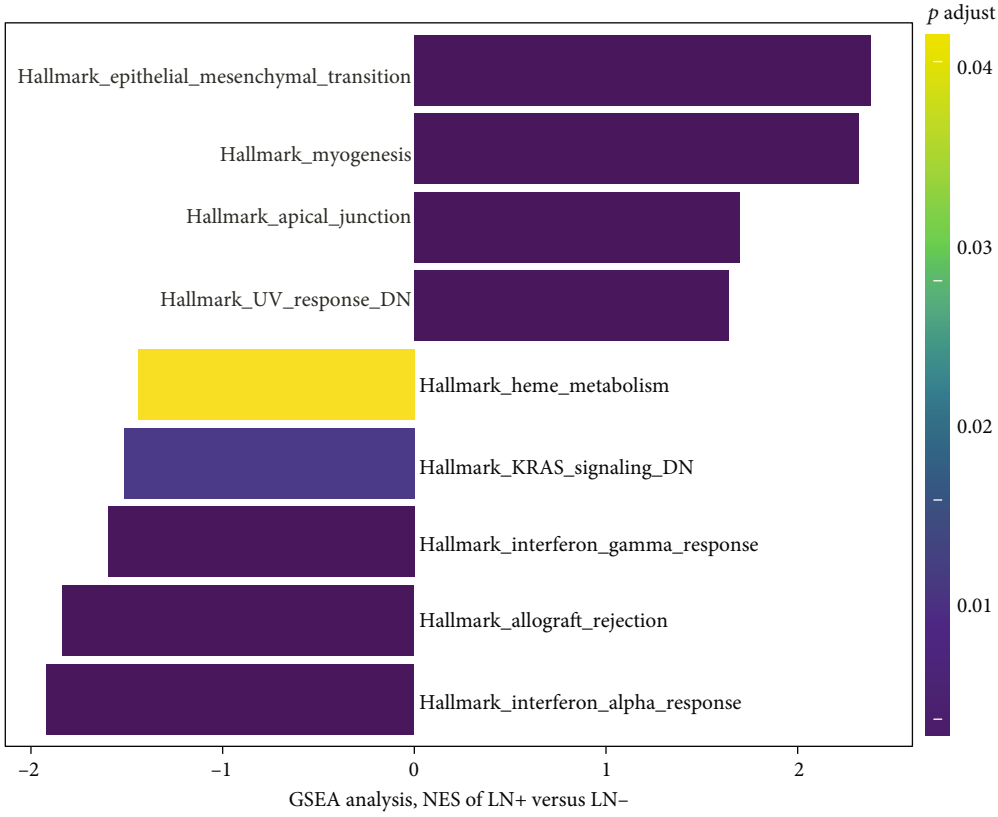
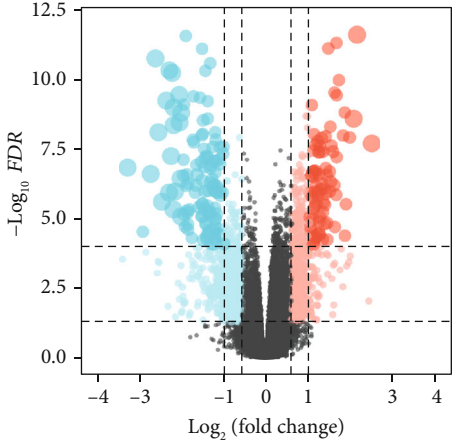
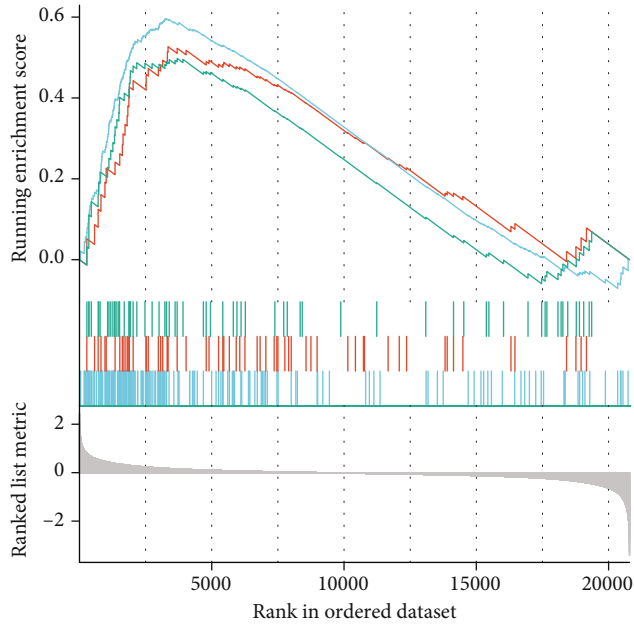
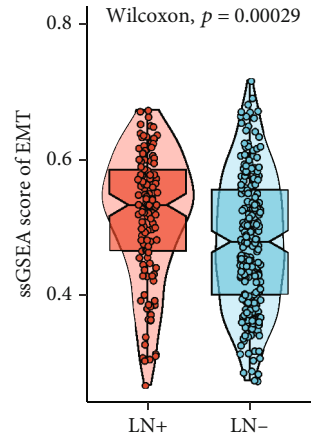


FIGURE 2: Continued.



— Gotzmann_epithelial_to_mesenchymal_transition_up
— Hallmark_epithelial_mesenchymal_transition
— Jechlinger_epithelial_to_mesenchymal_transition_up

(c)



(d)



(e)

FIGURE 2: Continued.

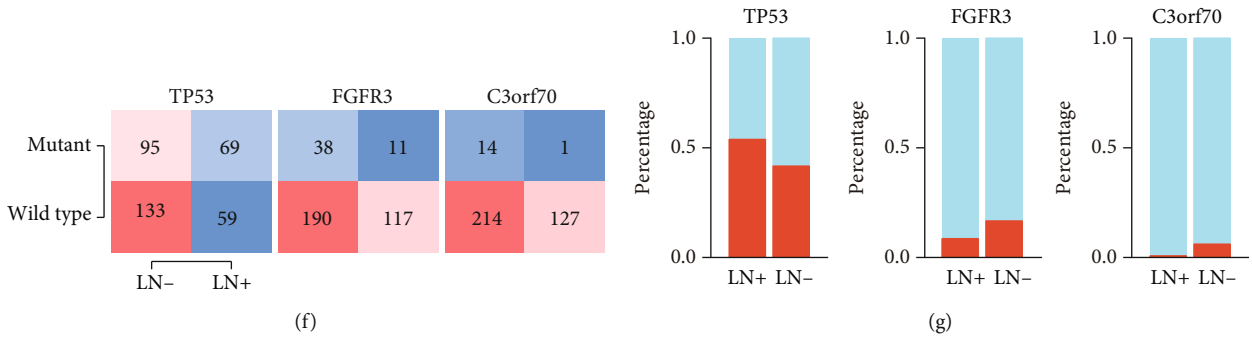


FIGURE 2: Overview of the molecular differences between LN+ and LN- tumours in the TCGA-BLCA cohort. (a) Volcano plot for differentially expressed genes. The cut-offs were set as follows: $\log_2(\text{foldchange}) > 1$, false discovery rate (FDR) < 0.05 . Red indicated upregulated genes; blue indicated downregulated genes. (b) Bar plot demonstrated the enrichment of hallmark gene sets. The shades of the bar represented the adjustment p value. (c) GSEA showed the high regulation of EMT-related pathways in LN+ tumours compared with LN- tumours. (d) Boxplot showed a significantly higher ssGSEA score of hallmark EMT gene set in LN+ tumours as compared to LN- tumours. (e) Oncoprint showed distribution of SMGs identified by MutSigCV in BLCA. The genetic alteration types were indicated in the waterfall plot annotation. The number on the left and right bar plots showed the mutation frequency of each gene. The LN metastasis status was shown as patient annotations in the bottom. (f) The heat map showed the difference in the number of patients with mutation of TP53, FGFR3, and C3orf70 between LN+ and LN- tumours. (g) The stacked bar plot showed the difference in the percentage of patients with mutation of TP53, FGFR3, and C3orf70 between LN+ and LN- tumours.

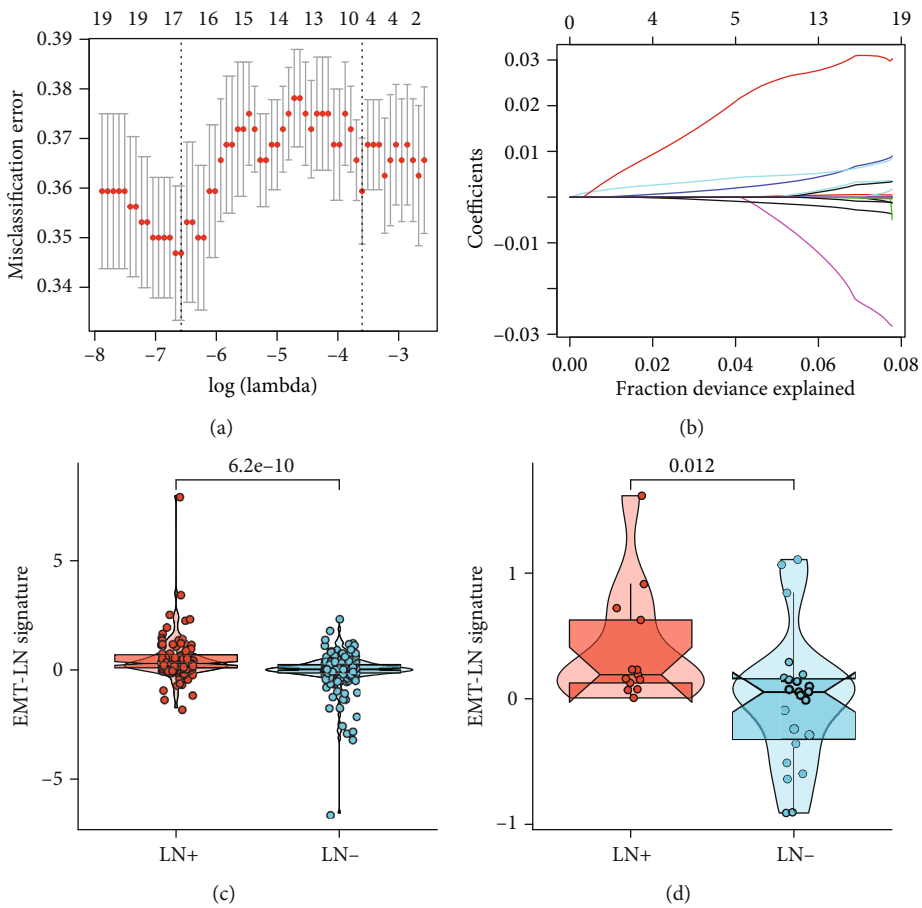


FIGURE 3: Feature selection using LASSO logistic regression model. (a) Tuning parameter (λ) selection 10-fold cross-validation error curve. The misclassification error was plotted vs. $\log(\lambda)$. The dotted vertical lines were drawn at the optimal values by the minimum criteria and the 1-SE criteria. (b) LASSO coefficient profiles of the 19 candidate EMT-related genes. A coefficient profile plot was produced against the $\log(\lambda)$ sequence. A vertical line was drawn at the value selected by 10-fold cross-validation, where the optimal λ resulted in 17 nonzero coefficients. (c, d) Difference in the EMT-LN signature risk score between LN+ and LN- tumours in the training dataset (c) and testing dataset (d) of the TCGA-BLCA cohort.

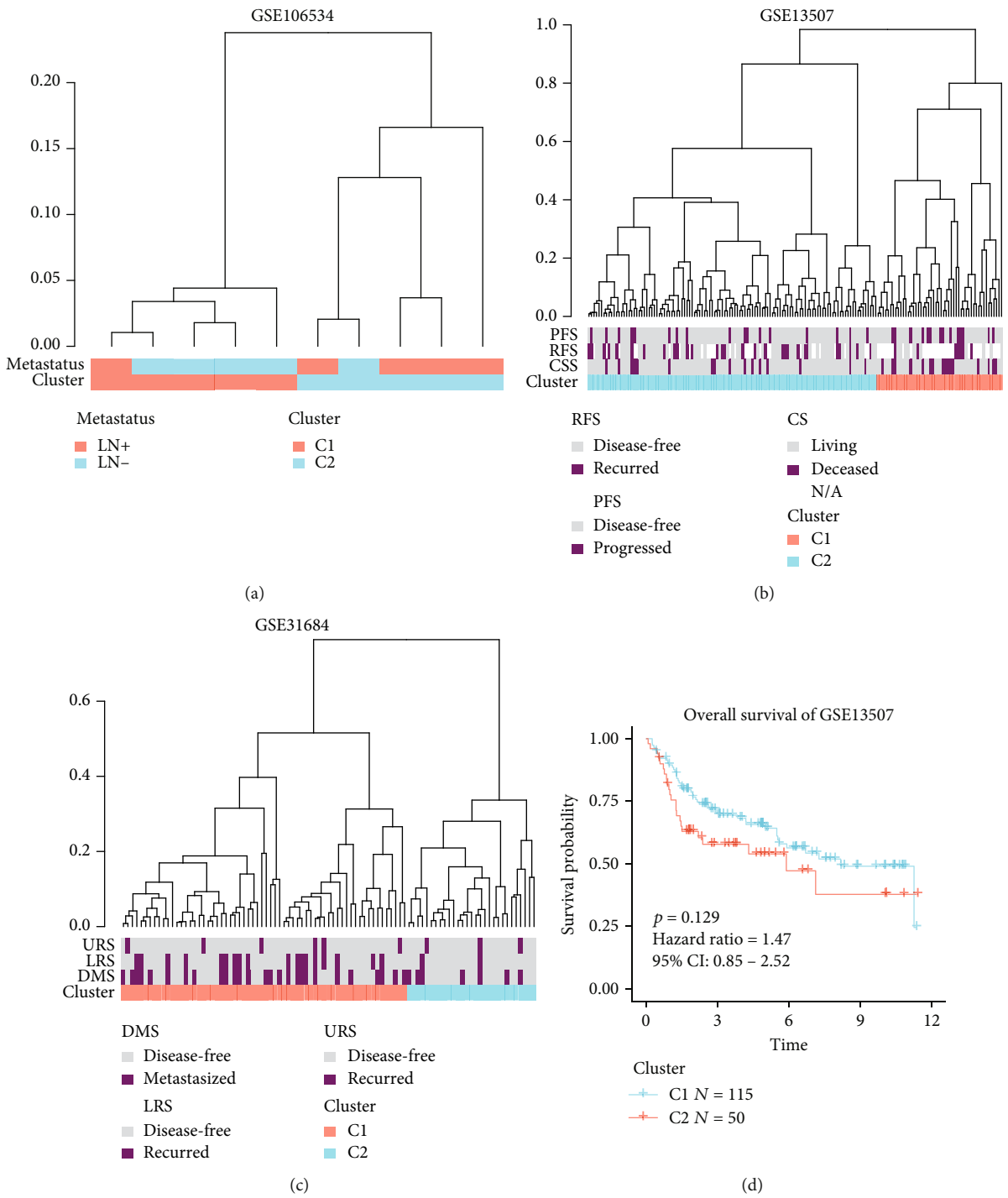


FIGURE 4: Continued.

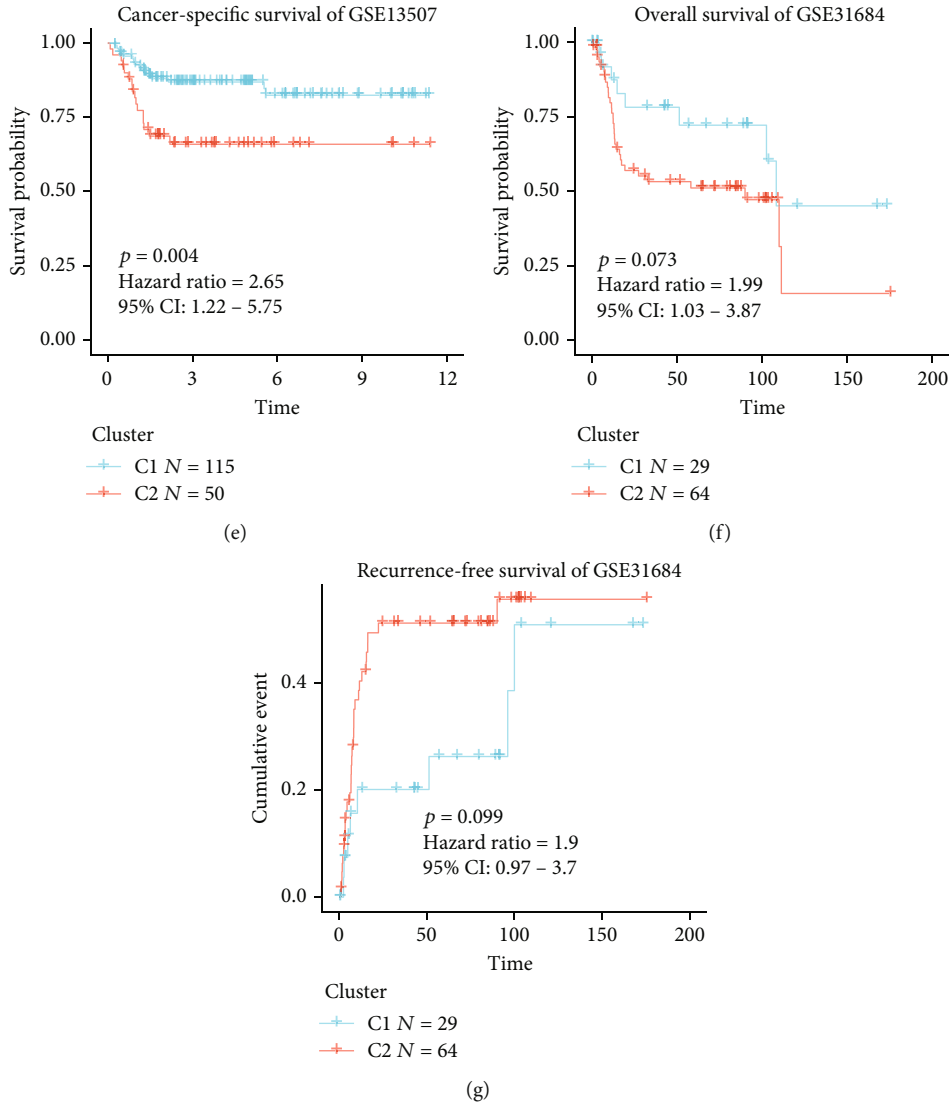


FIGURE 4: Supervised hierarchical clustering for EMT-LN signature. (a) Dendrogram showed that two clusters created by supervised hierarchical clustering could significantly distinguish LN metastasis status in the GSE106534 cohort. (b) Dendrogram showed that two clusters created by supervised hierarchical clustering were strikingly associated with OS, RFS, and CSS in the GSE13507 cohort. (c) Dendrogram showed that two clusters created by supervised hierarchical clustering were strikingly associated with DMS, LRS, and URS in the GSE31684 cohort. (d, e) Kaplan-Meier survival curves showed the difference in OS (log-rank test, $p = 0.129$, d) and CSS (log-rank test, $p = 0.004$, e) between LN+ and LN- tumours in the GSE13507 cohort. (f, g) Kaplan-Meier survival curves showed the difference in OS (log-rank test, $p = 0.073$, f) and CSS (log-rank test, $p = 0.099$, g) between LN+ and LN- tumours in the GSE31684 cohort.

4. Discussion

Bladder cancer (BLCA) is a lethal disease that little progress has been made in the past decades to improve the clinical outcome until the development of immune checkpoint inhibitors (ICIs) targeting immunotherapy [44]. Moreover, many clinical trials revealed that the patients who received ICIs immunotherapy did not all respond to it and non-responders will finally recur or progress to upstaging or even lymph node (LN) metastasis [45, 46]. A considerable amount of research demonstrated that BLCA patients with LN metastasis exhibited a poorer prognosis compared with patients without lymphatic spread [11]. Several lines of clinical evidence supported the idea that LN metastasis played signifi-

cant roles in systemic dissemination of cancer cells and has been recognized as a robust independent prognostic factor for clinical outcomes of BLCA patients [47]. Furthermore, the addition of pelvic lymphadenectomy (PLND) to radical cystectomy (RC) has also been reported to improve oncological outcomes even in patients without LN metastasis preoperative and would give a privilege for adjuvant therapy (AC) [48, 49]. These findings not only emphasized paramount importance of LN metastasis in BLCA development but also outlined the necessity of standardization of systemic therapy including PLND in the treatment of BLCA. Recently, neoadjuvant therapy (NAC) was widely used to improve the survival rate of BLCA patients. But some study demonstrated that preoperative utility of NAC in BLCA patients with occult

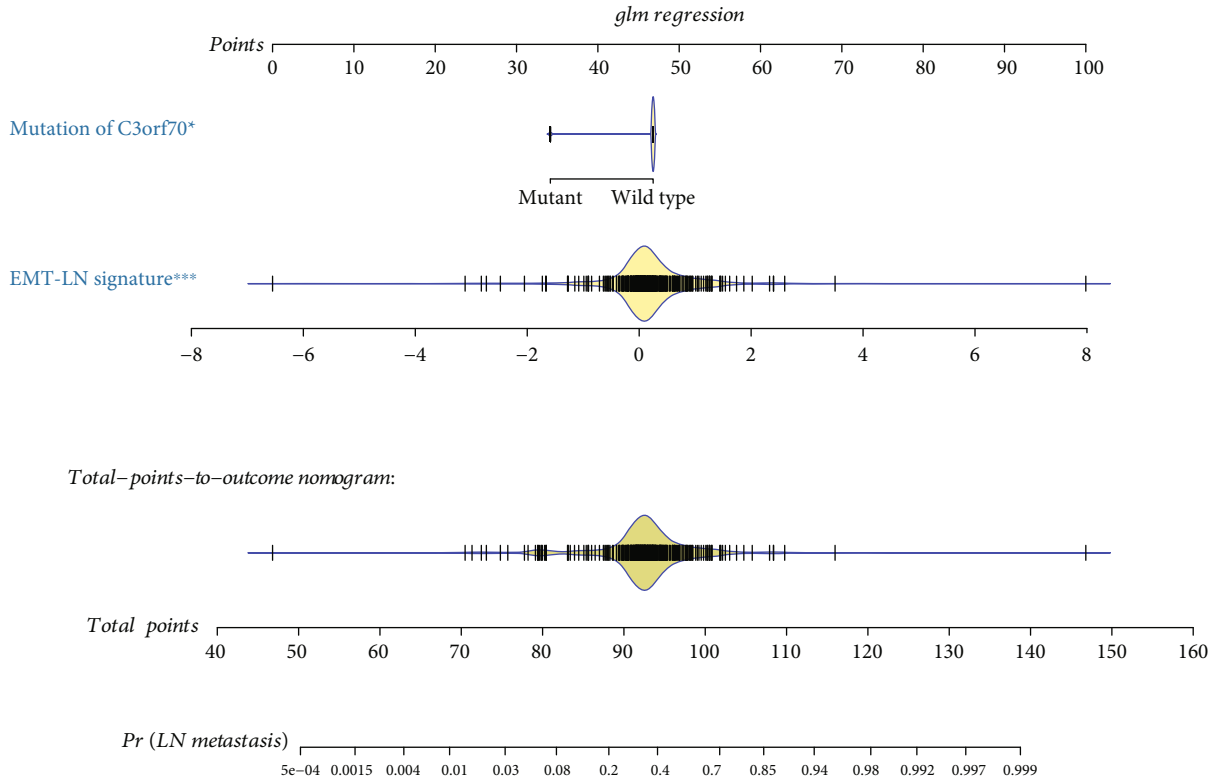


FIGURE 5: Development of preoperative EMT-LN nomogram. By incorporating the EMT-LN signature and genomic mutation of C3orf70, the EMT-LN nomogram was built in the training dataset of the TCGA-BLCA cohort.

LN metastases would induce a poor prognosis compared with patients treated with AC [20], which indicated the complexity of BLCA treatment and the extreme essence to know the LN metastasis status when applying distinct regimens.

Several histopathologic parameters are available to predict LN metastasis postoperatively, but none preoperatively. To date, medical imaging techniques have made a great contribution in preoperative diagnosis of LN metastasis, but it also exhibits substantial false positives that LN metastasis is unexpectedly found in 25% of BLCA patients undergoing RC and PLND who are diagnosed as stage N0 preoperatively by computed tomography (CT) or magnetic resonance imaging (MRI) [50]. These findings indicate that we could not just depend on the imaging techniques to predict LN metastasis.

Over the past couple of decades, researchers have paid increasing interest to molecular markers at achieving better prognostic stratification of cancer patients. Some groups even found molecular markers such as UPII and MUC7 to detect LN metastasis and prognosis in the setting of a positive marker with a negative lymph node at pathologic examination [28, 51]. However, some argued that qRT-PCR is over-sensitive to achieve positive results even in the absence of viable cancer cells [52], while others proposed that protein detection including immunohistochemistry (IHC) is not capable of reliably detecting somatic mutations. In addition, Malmstrom et al. even reported the inconsistency of IHC and quantitative real-time PCR (qRT-PCR) that only 50% of positive p53 patients with PCR had detectable protein expression at IHC [53]. In addition, with the development

of new technologies such as microarray and RNA sequencing, we could easily evaluate multiple molecules at the same time instead of detecting a single marker using qRT-PCR, which broadens our eyes in biomarker finding and improves the sensitivity of previous assays [54].

Epithelial-mesenchymal transition (EMT), which was first observed during embryonic development [55], now has been mainly recognized as an essential process in neoplastic progression [56, 57]. With the genetic and epigenetic alteration, neoplastic cells induced oncogenic EMT to favour clonal outgrowth and localized tumours development. Many reports also demonstrated that induction of EMT was of vital significance for LN metastatic in different types of cancer, including gastric cancer and head and neck cancer [58, 59]. Moreover, our previous study revealed that EMT is essential for the histology subtype transition from NMIBC to MIBC and strikingly correlated with survival in several independent cohorts [31]. In the present study, we found that EMT was significantly enriched in LN+ metastasis tumours compared with LN- tumours. Due to the aforementioned aspects, the importance of EMT in LN metastasis in BLCA is partly confirmed and we aimed to establish a diagnostic nomogram incorporating EMT for preoperative prediction of LN metastasis in BLCA patients.

For the construction of the EMT-LN signature, we firstly screened out 352 DEGs between LN+ and LN- tumour samples, merging them with genes from hallmark gene set to obtain 19 EMT-related candidate genes. Then, we performed LASSO logistic regression analysis to establish a 17-gene

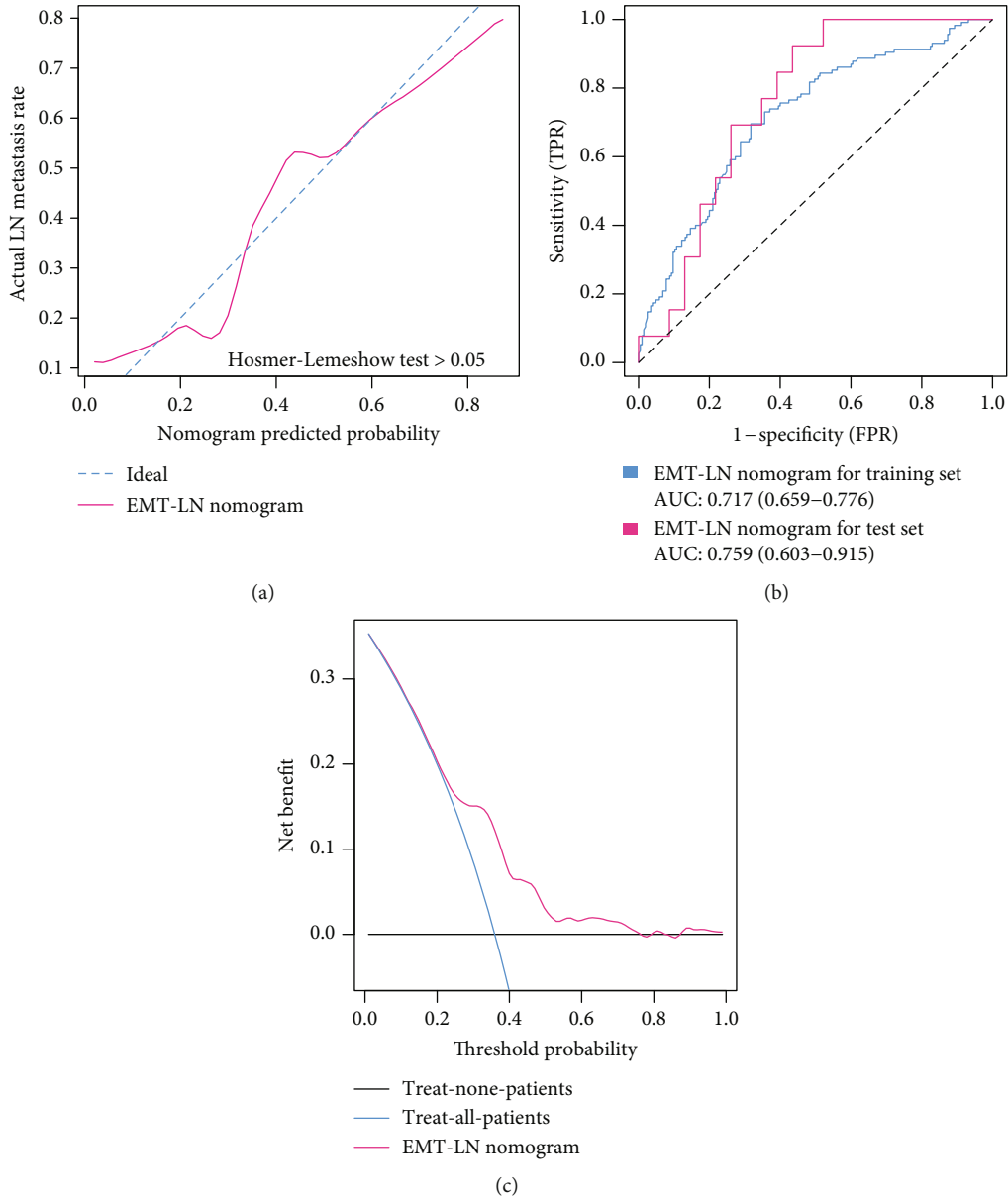


FIGURE 6: Clinical utility of the EMT-LN nomogram. (a) Calibration curve of the EMT-LN nomogram in the training dataset of the TCGA-BLCA cohort, which depicts the calibration of the fitted model between the predicted risk of LN metastasis and actual LN metastasis rate. The x -axis represents the predicted LN metastasis risk, and the y -axis represents the actual LN metastasis rate. The pink solid line represents the performance of the EMT-LN nomogram, of which a closer fit to the diagonal dotted blue line represents an ideal prediction. The difference in the two models was measured with the Hosmer-Lemeshow test. (b) ROC curves showed the prediction accuracy of the EMT-LN signature in prediction of the LN metastasis in training and testing datasets of the TCGA-BLCA cohort. (c) Decision curve analysis (DCA) for the EMT-LN nomogram. The y -axis measures the net benefit. The pink line represents the EMT-LN nomogram, the blue line represents the assumption that all patients have LN metastases, and the black line on the bottom represents the assumption that no patients have LN metastases.

EMT-related signature for predicting LN metastasis status by shrinking the regression coefficients for dimension reduction. The LASSO algorithm could not only reduce noise or redundant features but also combine all selected features and assembles a single signature, i.e., marker panels instead of single predictor selection based on strength of univariable regression analysis. Recently, multiple biomarker panels have been merged as novel biomarkers for outcome prediction and are proved to be more effective than single biomarkers,

owing to the high disease heterogeneity [60]. The Oncotype DX genomic testing is the first clinically used and validated multigene assay, which can quantify the likelihood of breast cancer recurrence [61, 62]. Chen et al. demonstrated a 10-gene biomarker panel based on the consolidation of data derived from transcriptomics and proteomics to predict the occurrence of BLCA with diagnostic sensitivity of 79% and specificity of 79% in a multicenter cohort [63]. Recently, our group even constructed an immune-related lncRNA

panel to predict the immunotherapeutic efficiency in BLCA [64]. Similarly, the EMT-LN signature established in the recent study showed a significantly different distribution in LN+ and LN- tumours in TCGA-BLCA training and testing cohorts. The EMT-LN signature was also presented as a prognostic independent factor for prognosis in the TCGA-BLCA cohort. In addition, the EMT-LN signature could also allow us to distinguish LN metastasis status through supervised hierarchical clustering and was highly associated with patients' outcomes to some extent in several independent GEO cohorts. In total, the EMT-LN signature was a noninvasive factor for us to predict the LN metastasis preoperatively.

The landscape of somatic mutation was evaluated across LN+ and LN- tumours that mutation of TP53, FGFR3, and C3orf70 was differentially mutated between LN+ and LN- tumours. Moreover, multivariable logistic regression analysis demonstrated that only mutation of C3orf70 remained as a significant variable in the predictive model. Recently, C3orf70 was reported to be involved in neural and neurobehavioral development [65]. Ward et al. establish a 23-gene panel containing C3orf70 with utility for noninvasive diagnosis and risk stratification of BLCA [66]. There are no other reports about C3orf70, which push us to investigate its underlying role in BLCA development in the future. Then, we integrated the EMT-LN signature and C3orf70 mutation to generate a nomogram, which allowed the clinicians to predict LN metastasis risk in patients with BLCA preoperatively. The nomogram is an easy-to-use quantitative scoring tool that the individual will get total points by a plus single point of each feature within the nomogram, which is calculated through a specific algorithm by considering the contribution of features. Then, the total points could be transferred into the probability of the individual clinical ending event, such as LN metastasis. Next, we found that the EMT-LN nomogram was effective in discriminating the LN metastasis status with high accuracy in both TCGA-BLCA training (71.7%) and testing (75.9%) cohorts. The calibration curves with the Hosmer-Lemeshow test showed that the EMT-LN nomogram exhibited no significant difference with the ideal model. Moreover, DCA curves revealed that decisions based on the EMT-LN nomogram are superior compared with the treat-all patient scheme and the treat-none scheme, which give a large threshold probability and yielded more favourable clinical consequences. All of these suggested that our established nomogram could be of a high potential for clinical utility.

In summary, we establish an EMT-LN nomogram integrating an EMT-LN signature and C3orf70 mutation, which act as an easy-to-use tool to facilitate preoperative prediction of LN metastasis in BLCA individuals.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The data that support the findings of this study were derived from the following resources: TCGA GDC data portal (<https://portal.gdc.cancer.gov/>) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

RC and LY made substantial contributions to conception and design of the research. RC, LY, and BM integrated and analysed the data. RC, LY, and YT wrote the paper. RC, LY, BM, GW, and TY edited the manuscript and provided critical comments. All authors read and approved the final manuscript.

Acknowledgments

The authors are grateful for the invaluable support and useful discussions with colleagues in the Department of Urology. We thank the TCGA and GEO databases for the availability of the data. This work was supported by grants from the National Natural Science Foundation of China (Grant No. 81802550), China Postdoctoral Science Foundation (Grant No. 2019M660041), and Beijing Postdoctoral Research Foundation (Grant No. ZZ2019-04).

Supplementary Materials

Supplementary 1. Supplementary Table S1: differential expression analysis based on LN metastasis status by DESeq2.

Supplementary 2. Supplementary Table S2: GSEA based on preranked gene list calculated by DESeq2 for hallmark gene sets.

Supplementary 3. Supplementary Table S3: GSEA based on preranked gene list calculated by DESeq2 for EMT-related gene sets.

Supplementary 4. Supplementary Table S4: significantly mutated genes identified by MutSigCV.

Supplementary 5. Supplementary Table S5: differentially mutated genes based on LN metastasis status.

Supplementary 6. Supplementary Table S6: selected features with nonzero coefficients in the LASSO logistic regression model.

Supplementary 7. Supplementary Table S7: multivariable logistic regression analysis for preoperative features.

Supplementary 8. Figure S1: the design and workflow of the study.

Supplementary 9. Figure S2: ROC curves showed the prediction accuracy of EMT-LN signature in prediction of the LN metastasis in GSE13507, GSE31684, and GSE106534 cohorts.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in

- 185 countries,” *CA: a Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] O. Sanli, J. Dobruch, M. A. Knowles et al., “Bladder cancer,” *Nature reviews Disease primers*, vol. 3, no. 1, article 17022, 2017.
- [3] S. L. Woldu, A. Bagrodia, and Y. Lotan, “Guideline of guidelines: non-muscle-invasive bladder cancer,” *BJU international*, vol. 119, no. 3, pp. 371–380, 2017.
- [4] H. M. Fritsche, M. Burger, R. S. Svatek et al., “Characteristics and outcomes of patients with clinical T1 grade 3 urothelial carcinoma treated with radical cystectomy: results from an international cohort,” *European Urology*, vol. 57, no. 2, pp. 300–309, 2010.
- [5] G. S. Kulkarni, T. Hermanns, Y. Wei et al., “Propensity score analysis of radical cystectomy versus bladder-sparing trimodal therapy in the setting of a multidisciplinary bladder cancer clinic,” *Journal of clinical oncology*, vol. 35, no. 20, pp. 2299–2305, 2017.
- [6] M. I. Milowsky, R. B. Rumble, C. M. Booth et al., “Guideline on muscle-invasive and metastatic bladder cancer (European Association of Urology guideline): American Society of Clinical Oncology clinical practice guideline endorsement,” *Journal of clinical oncology*, vol. 34, no. 16, pp. 1945–1952, 2016.
- [7] E. E. Vokes, N. Ready, E. Felip et al., “Nivolumab versus docetaxel in previously treated advanced non-small-cell lung cancer (CheckMate 017 and CheckMate 057): 3-year update and outcomes in patients with liver metastases,” *Annals of oncology*, vol. 29, no. 4, pp. 959–965, 2018.
- [8] P. Sharma, M. Retz, A. Siefker-Radtke et al., “Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial,” *The Lancet Oncology*, vol. 18, no. 3, pp. 312–322, 2017.
- [9] J. Larkin, D. Minor, S. D’Angelo et al., “Overall survival in patients with advanced melanoma who received nivolumab versus investigator’s choice chemotherapy in CheckMate 037: a randomized, controlled, open-label phase III trial,” *Journal of clinical oncology*, vol. 36, no. 4, pp. 383–390, 2018.
- [10] J. A. Witjes, H. M. Bruins, R. Cathomas et al., “European Association of Urology guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2020 guidelines,” *European Urology*, 2020.
- [11] J. P. Stein, J. Cai, S. Groshen, and D. G. Skinner, “Risk factors for patients with pelvic lymph node metastases following radical cystectomy with en bloc pelvic lymphadenectomy: concept of lymph node density,” *The Journal of urology*, vol. 170, no. 1, pp. 35–41, 2003.
- [12] A. Karl, P. R. Carroll, J. E. Gschwend et al., “The impact of lymphadenectomy and lymph node metastasis on the outcomes of radical cystectomy for bladder cancer,” *European Urology*, vol. 55, no. 4, pp. 826–835, 2009.
- [13] B. R. Konety, S. A. Joslyn, and M. A. O’Donnell, “Extent of pelvic lymphadenectomy and its impact on outcome in patients diagnosed with bladder cancer: analysis of data from the Surveillance, Epidemiology and End Results program data base,” *The Journal of urology*, vol. 169, no. 3, pp. 946–950, 2003.
- [14] W. Kassouf, R. S. Svatek, S. F. Shariat et al., “Critical analysis and validation of lymph node density as prognostic variable in urothelial carcinoma of bladder,” *Urologic Oncology*, vol. 31, no. 4, pp. 480–486, 2013.
- [15] M. Perera, S. McGrath, S. Sengupta, J. Crozier, D. Bolton, and N. Lawrentschuk, “Pelvic lymph node dissection during radical cystectomy for muscle-invasive bladder cancer,” *Nature reviews Urology*, vol. 15, no. 11, pp. 686–692, 2018.
- [16] J. B. Jensen, B. P. Ulhoi, and K. M. Jensen, “Extended versus limited lymph node dissection in radical cystectomy: impact on recurrence pattern and survival,” *International journal of urology*, vol. 19, no. 1, pp. 39–47, 2012.
- [17] G. Simone, R. Papalia, M. Ferriero et al., “Stage-specific impact of extended versus standard pelvic lymph node dissection in radical cystectomy,” *International journal of urology*, vol. 20, no. 4, pp. 390–397, 2013.
- [18] J. E. Rosenberg, “Current status of neoadjuvant and adjuvant chemotherapy for muscle-invasive bladder cancer,” *Expert review of anticancer therapy*, vol. 7, pp. 1729–1736, 2014.
- [19] C. C. Peyton, D. Tang, R. R. Reich et al., “Downstaging and survival outcomes associated with neoadjuvant chemotherapy regimens among patients treated with cystectomy for muscle-invasive bladder cancer,” *JAMA oncology*, vol. 4, no. 11, pp. 1535–1542, 2018.
- [20] E. K. Cha, J. P. Sfakianos, R. Sukhu, A. M. Yee, D. D. Sjoberg, and B. H. Bochner, “Poor prognosis of bladder cancer patients with occult lymph node metastases treated with neoadjuvant chemotherapy,” *BJU international*, vol. 122, no. 4, pp. 627–632, 2018.
- [21] F. Liedberg, G. Chebil, T. Davidsson, S. Gudjonsson, and W. Mansson, “Intraoperative sentinel node detection improves nodal staging in invasive bladder cancer,” *The Journal of urology*, vol. 175, no. 1, pp. 84–88, 2006, discussion 88–89.
- [22] B. Roth, M. P. Wissmeyer, P. Zehnder et al., “A new multimodality technique accurately maps the primary lymphatic landing sites of the bladder,” *European Urology*, vol. 57, no. 2, pp. 205–211, 2010.
- [23] C. J. McMahon, N. M. Rofsky, and I. Pedrosa, “Lymphatic metastases from pelvic tumors: anatomic classification, characterization, and staging,” *Radiology*, vol. 254, no. 1, pp. 31–46, 2010.
- [24] A. Fleischmann, G. N. Thalmann, R. Markwalder, and U. E. Studer, “Extracapsular extension of pelvic lymph node metastases from urothelial carcinoma of the bladder is an independent prognostic factor,” *Journal of clinical oncology*, vol. 23, no. 10, pp. 2358–2365, 2005.
- [25] M. L. Paik, M. J. Scolieri, S. L. Brown, J. P. Spirnak, and M. I. Resnick, “Limitations of computerized tomography in staging invasive bladder cancer before radical cystectomy,” *The Journal of urology*, vol. 163, no. 6, pp. 1693–1696, 2000.
- [26] L. Evangelista, A. Guttilla, F. Zattoni, P. C. Muzzio, and F. Zattoni, “Utility of choline positron emission tomography/computed tomography for lymph node involvement identification in intermediate- to high-risk prostate cancer: a systematic literature review and meta-analysis,” *European Urology*, vol. 63, no. 6, pp. 1040–1048, 2013.
- [27] H. L. Copp, J. L. Chin, M. Conaway, and D. Theodorescu, “Prospective evaluation of the prognostic relevance of molecular staging for urothelial carcinoma,” *Cancer*, vol. 107, no. 1, pp. 60–66, 2006.
- [28] M. Retz, J. Lehmann, C. Szysnik et al., “Detection of occult tumor cells in lymph nodes from bladder cancer patients by MUC7 nested RT-PCR,” *European Urology*, vol. 45, no. 3, pp. 314–319, 2004.
- [29] M. Birkhahn, A. P. Mitra, and R. J. Cote, “Molecular markers for bladder cancer: the road to a multimarker approach,”

- Expert review of anticancer therapy*, vol. 7, pp. 1717–1727, 2014.
- [30] K. E. van Kessel, H. J. van de Werken, I. Lurkin, A. C. Ziel-van der Made, E. C. Zwarthoff, and J. L. Boormans, “A reported 20-gene expression signature to predict lymph node-positive disease at radical cystectomy for muscle-invasive bladder cancer is clinically not applicable,” *PLoS One*, vol. 12, no. 3, article e0174039, 2017.
- [31] R. Cao, L. Yuan, B. Ma, G. Wang, W. Qiu, and Y. Tian, “An EMT-related gene signature for the prognosis of human bladder cancer,” *Journal of cellular and molecular medicine*, vol. 24, no. 1, pp. 605–617, 2020.
- [32] G. P. Wagner, K. Kin, and V. J. Lynch, “Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples,” *Theorie in den Biowissenschaften*, vol. 131, no. 4, pp. 281–285, 2012.
- [33] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, “affy-analysis of Affymetrix GeneChip data at the probe level,” *Bioinformatics*, vol. 20, no. 3, pp. 307–315, 2004.
- [34] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, “clusterProfiler: an R package for comparing biological themes among gene clusters,” *OmicS: a journal of integrative biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [35] A. Subramanian, P. Tamayo, V. K. Mootha et al., “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [36] A. Colaprico, T. C. Silva, C. Olsen et al., “TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data,” *Nucleic Acids Research*, vol. 44, no. 8, article e71, 2016.
- [37] M. S. Lawrence, P. Stojanov, P. Polak et al., “Mutational heterogeneity in cancer and the search for new cancer-associated genes,” *Nature*, vol. 499, no. 7457, pp. 214–218, 2013.
- [38] Z. Gu, R. Eils, and M. Schlesner, “Complex heatmaps reveal patterns and correlations in multidimensional genomic data,” *Bioinformatics*, vol. 32, no. 18, pp. 2847–2849, 2016.
- [39] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [40] J. M. Bland and D. G. Altman, “Survival probabilities (the Kaplan-Meier method),” *BMJ*, vol. 317, no. 7172, pp. 1572–1580, 1998.
- [41] A. Mayakonda, D. C. Lin, Y. Assenov, C. Plass, and H. P. Koefler, “Maftools: efficient and comprehensive analysis of somatic variants in cancer,” *Genome Research*, vol. 28, no. 11, pp. 1747–1756, 2018.
- [42] A. Iasonos, D. Schrag, G. V. Raj, and K. S. Panageas, “How to build and interpret a nomogram for cancer prognosis,” *Journal of Clinical Oncology*, vol. 26, no. 8, pp. 1364–1370, 2008.
- [43] A. G. Robertson, J. Kim, H. al-Ahmadie et al., “Comprehensive molecular characterization of muscle-invasive bladder cancer,” *Cell*, vol. 171, no. 3, pp. 540–556.e25, 2017.
- [44] B. C. Holland, A. Sood, K. Delfino et al., “Age and sex have no impact on expression levels of markers of immune cell infiltration and immune checkpoint pathways in patients with muscle-invasive urothelial carcinoma of the bladder treated with radical cystectomy,” *Cancer immunology, immunotherapy*, vol. 68, no. 6, pp. 991–997, 2019.
- [45] W. Obara, I. Hara, Y. Kato et al., “Immunotherapy with cancer peptides in combination with intravesical bacillus Calmette-Guerin for patients with non-muscle invasive bladder cancer,” *Cancer immunology, immunotherapy*, vol. 67, no. 9, pp. 1371–1380, 2018.
- [46] M. Gupta, M. Kates, and T. J. Bivalacqua, “Immunotherapy in nonmuscle invasive bladder cancer: current and emerging treatments,” *Current opinion in oncology*, vol. 31, no. 3, pp. 183–187, 2019.
- [47] K. Kawada and M. M. Taketo, “Significance and mechanism of lymph node metastasis in cancer progression,” *Cancer research*, vol. 71, no. 4, pp. 1214–1218, 2011.
- [48] H. W. Herr, B. H. Bochner, G. Dalbagni, S. M. Donat, V. E. Reuter, and D. F. Bajorin, “Impact of the number of lymph nodes retrieved on outcome in patients with muscle invasive bladder cancer,” *The Journal of urology*, vol. 167, no. 3, pp. 1295–1298, 2002.
- [49] W. Kassouf, P. K. Agarwal, H. B. Grossman et al., “Outcome of patients with bladder cancer with pN+ disease after preoperative chemotherapy and radical cystectomy,” *Urology*, vol. 73, no. 1, pp. 147–152, 2009.
- [50] S. F. Shariat, G. S. Palapattu, P. I. Karakiewicz et al., “Discrepancy between clinical and pathologic stage: impact on prognosis after radical cystectomy,” *European Urology*, vol. 51, no. 1, pp. 137–151, 2007, discussion 149–151.
- [51] X. Wu, Y. Takehi, Y. Zeng, R. Taoka, H. Tsunemori, and M. Inui, “Uroplakin II as a promising marker for molecular diagnosis of nodal metastases from bladder cancer: comparison with cytokeratin 20,” *The Journal of urology*, vol. 174, no. 6, pp. 2138–2143, 2005.
- [52] S. A. Keene and M. J. Demeure, “The clinical significance of micrometastases and molecular metastases,” *Surgery*, vol. 129, no. 1, pp. 1–5, 2001.
- [53] P. U. Malmstrom, Z. P. Ren, A. Sherif, M. de la Torre, K. Wester, and M. Thorn, “Early metastatic progression of bladder carcinoma: molecular profile of primary tumor and sentinel lymph node,” *The Journal of urology*, vol. 168, no. 5, pp. 2240–2244, 2002.
- [54] U. E. Gibson, C. A. Heid, and P. M. Williams, “A novel method for real time quantitative RT-PCR,” *Genome Research*, vol. 6, no. 10, pp. 995–1001, 1996.
- [55] P. Savagner, B. Boyer, A. M. Valles, J. Jouanneau, and J. P. Thiery, “Modulations of the epithelial phenotype during embryogenesis and cancer progression,” *Cancer treatment and research*, vol. 71, pp. 229–249, 1994.
- [56] D. Tarin, E. W. Thompson, and D. F. Newgreen, “The fallacy of epithelial mesenchymal transition in neoplasia,” *Cancer research*, vol. 65, no. 14, pp. 5996–6001, 2005, discussion 6000–5991.
- [57] J. P. Thiery, “Epithelial-mesenchymal transitions in tumour progression,” *Nature Reviews. Cancer*, vol. 2, no. 6, pp. 442–454, 2002.
- [58] W. Y. Lee, D. Y. Shin, H. J. Kim, Y. H. Ko, S. Kim, and H. S. Jeong, “Prognostic significance of epithelial-mesenchymal transition of extracapsular spread tumors in lymph node metastases of head and neck cancer,” *Annals of surgical oncology*, vol. 21, no. 6, pp. 1904–1911, 2014.
- [59] K. Okubo, Y. Uenosono, T. Arigami et al., “Clinical significance of altering epithelial-mesenchymal transition in metastatic lymph nodes of gastric cancer,” *Gastric cancer*, vol. 20, no. 5, pp. 802–810, 2017.

- [60] M. Frantzi, A. Latosinska, L. Fluhe et al., “Developing proteomic biomarkers for bladder cancer: towards clinical application,” *Nature reviews Urology*, vol. 12, no. 6, pp. 317–330, 2015.
- [61] T. P. McVeigh, L. M. Hughes, N. Miller et al., “The impact of Oncotype DX testing on breast cancer management and chemotherapy prescribing patterns in a tertiary referral centre,” *European Journal of Cancer*, vol. 50, no. 16, pp. 2763–2770, 2014.
- [62] T. P. McVeigh and M. J. Kerin, “Clinical use of the Oncotype DX genomic test to guide treatment decisions for patients with invasive breast cancer,” *Breast Cancer*, vol. 9, pp. 393–400, 2017.
- [63] L. M. Chen, M. Chang, Y. Dai et al., “External validation of a multiplex urinary protein panel for the detection of bladder cancer in a multicenter cohort,” *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, vol. 23, no. 9, pp. 1804–1812, 2014.
- [64] R. Cao, L. Yuan, B. Ma, G. Wang, and Y. Tian, “Immune-related long non-coding RNA signature identified prognosis and immunotherapeutic efficiency in bladder cancer (BLCA),” *Cancer cell international*, vol. 20, no. 1, p. 276, 2020.
- [65] Y. Ashikawa, T. Shiromizu, K. Miura et al., “C3orf70 is involved in neural and neurobehavioral development,” *Pharmaceuticals*, vol. 12, no. 4, p. 156, 2019.
- [66] D. G. Ward, N. S. Gordon, R. H. Boucher et al., “Targeted deep sequencing of urothelial bladder cancers and associated urinary DNA: a 23-gene panel with utility for non-invasive diagnosis and risk stratification,” *BJU international*, vol. 124, no. 3, pp. 532–544, 2019.