



A Qualitative Transcriptional Signature for Predicting Recurrence Risk of Stage I–III Bladder Cancer Patients After Surgical Resection

Yawei Li¹, Huarong Zhang¹, You Guo², Hao Cai², Xiangyu Li¹, Jun He¹, Hung-Ming Lai¹, Qingzhou Guan¹, Xianlong Wang^{1,3*} and Zheng Guo^{1,3*}

¹ Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Department of Bioinformatics, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China, ² Medical Big Data and Bioinformatics Research Centre, First Affiliated Hospital of Gannan Medical University, Ganzhou, China, ³ Key Laboratory of Medical Bioinformatics, Fujian Medical University, Fuzhou, China

OPEN ACCESS

Edited by:

Fabio Grizzi,
Humanitas Research Hospital, Italy

Reviewed by:

Naokazu Ibuki,
Osaka Medical College, Japan
Riccardo Autorino,
Virginia Commonwealth University,
United States

*Correspondence:

Zheng Guo
guoz@ems.hrbmu.edu.cn

Specialty section:

This article was submitted to
Genitourinary Oncology,
a section of the journal
Frontiers in Oncology

Received: 15 November 2018

Accepted: 25 June 2019

Published: 10 July 2019

Citation:

Li Y, Zhang H, Guo Y, Cai H, Li X,
He J, Lai H-M, Guan Q, Wang X and
Guo Z (2019) A Qualitative
Transcriptional Signature for Predicting
Recurrence Risk of Stage I–III Bladder
Cancer Patients After Surgical
Resection. *Front. Oncol.* 9:629.
doi: 10.3389/fonc.2019.00629

Background: Previously reported transcriptional signatures for predicting the prognosis of stage I–III bladder cancer (BLCA) patients after surgical resection are commonly based on risk scores summarized from quantitative measurements of gene expression levels, which are highly sensitive to the measurement variation and sample quality and thus hardly applicable under clinical settings. It is necessary to develop a signature which can robustly predict recurrence risk of BLCA patients after surgical resection.

Methods: The signature is developed based on the within-sample relative expression orderings (REOs) of genes, which are qualitative transcriptional characteristics of the samples.

Results: A signature consisting of 12 gene pairs (12-GPS) was identified in training data with 158 samples. In the first validation dataset with 114 samples, the low-risk group of 54 patients had a significantly better overall survival than the high-risk group of 60 patients (HR = 3.59, 95% CI: 1.34~9.62, $p = 6.61 \times 10^{-03}$). The signature was also validated in the second validation dataset with 57 samples (HR = 2.75×10^{08} , 95% CI: 0~Inf, $p = 0.05$). Comparison analysis showed that the transcriptional differences between the low- and high-risk groups were highly reproducible and significantly concordant with DNA methylation differences between the two groups.

Conclusions: The 12-GPS signature can robustly predict the recurrence risk of stage I–III BLCA patients after surgical resection. It can also aid the identification of reproducible transcriptional and epigenomic features characterizing BLCA metastasis.

Keywords: bladder cancer, micro-metastasis, qualitative transcriptional signature, differentially expressed genes, differentially methylated genes

INTRODUCTION

Bladder cancer (BLCA) is still a major health problem worldwide (1). For stage I–III BLCA patients without metastasis, transurethral resection of bladder tumor (stage I) or radical cystectomy (stage II–III) is the standard treatment, which can improve patients' survival. However, a study based on large samples reported that, after radical cystectomy, the 5 years mortality rate for stage I patients

and stage II-III patients are only about 18.4 and 36.2%, respectively (2, 3), leading to the poor prognosis of BLCA patients (4). An important reason is that occult micro-metastases contribute to high recurrence risk of patients (5). Currently, both imaging techniques (6–8) and pathological evaluation techniques even with an enlarging pathological evaluation (9) have relatively high false-negative rates for occult micro-metastases, which are likely responsible for most of the recurrence or mortality (6–8). Therefore, it is urgent to develop an accurate molecular signature for predicting potential micro-metastasis states of postoperative patients.

Many transcriptional signatures for predicting prognosis or recurrence risk of BLCA patients after curative surgery have been provided (10–13). However, a critical limitation of such quantitative signatures is that their applications are commonly based on risk scores calculated as quantitative summaries of the expression measurements of the signature genes, which are impractical for clinical applications due to large measurement batch effects (14, 15) and quality uncertainties of clinical samples (16–19). In contrast, the signatures based on the within-sample relative expression orderings (REOs) of genes are robust against experimental batch effects (20) and can be applied to individual disease samples assessed in different laboratories (14, 21–24). Besides, the REO-based signatures are highly robust against varied proportions of the tumor epithelial cell in tumor tissues sampled from different tumor locations of the same patient (18), partial RNA degradation during specimen storage and preparation (17) and amplification bias for minimum specimens even with about 15–25 cancer cells (19), which are common factors leading to failures of quantitative transcriptional signatures in clinical applications. Therefore, it is feasible to identify a REOs-based signature for predicting recurrence risk of stage I–III BLCA patients after surgical resection.

In this study, a REOs-based prognostic signature, composed of 12 gene pairs, was identified in training dataset, and this signature was validated in two independent datasets. Then, we used the BLCA samples from The Cancer Genome Atlas (TCGA), to analyze the genomic and epigenomic gene markers characterizing the two prognostic groups.

MATERIALS AND METHODS

Data Sources and Data Preprocessing

The TCGA data, measured by the Illumina-Hiseq platform, was downloaded from TCGA data portal website (<http://cancergenome.nih.gov/>). From the TCGA RNA-seq data, we extracted 158 samples of patients who did not receive chemotherapy, neoadjuvant or radiotherapy, denoted as BLCA158, for training the signature. The mRNA-seq profiles of level 3 Fragments Per Kilobase Million (FPKM) was extracted as the gene expression measurements for the dataset BLCA158. Another two datasets of BLCA samples profiled by different microarray platforms were downloaded from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). From the GSE31684 dataset measured by the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570), 57 samples of patients treated with curative surgery alone were extracted and used as the

TABLE 1 | Description of the datasets used in this study.

	Training datasets		Validation datasets	
	BLCA158		BLCA57	BLCA114
	TCGA		GSE31684	GSE32894
Platform	Illumina Hiseq-RNAseq		GPL570	GPL6974
Sample Size	158		57	114
Stage	0	0	5	0
	I	1	9	63
	II	60	12	43
	III	58	22	8
	IV	39	9	0

first validation dataset, denoted as BLCA57. From the GSE32894 dataset measured by the Illumina HumanHT platform GPL6974, we extracted 114 samples of stage I-III patients treated with curative surgery alone, excluding samples of patients treated with radiotherapy, as the second validation dataset, denoted as BLCA114. Data description is summarized in **Table 1**. Probe-set IDs were mapped to gene IDs using the corresponding platform files. If multiple probesets were mapped to the same gene, the expression value of the gene was summarized as the arithmetic mean of the values of the multiple probesets for the GSE31684 and GSE32894 datasets.

For the 158 samples of BLCA158 dataset, the DNA methylation beta-values of samples for 15,932 genes measured by the Infinium HumanMethylation450 platform was derived from the TCGA Web Portal.

Survival Analysis

The disease-free survival (DFS) was defined as the time from surgical resection to the date of tumor recurrence or distant metastasis, and overall survival (OS) was defined as the time from surgery to death. Survival curves of DFS and OS between distinct subgroups were estimated by the Kaplan-Meier method and compared with log-rank tests (25). The univariate Cox regression model was used to evaluate the correlation of expression values of genes and the REOs of gene pairs with patients' OS. The multivariate Cox regression model was used to evaluate the independent prognostic value of the signature after adjusting for clinical features including age, gender, and stage, which could be written as follow:

$$h(t, x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)$$

where $h(t, x)$ is the hazard function determined by a set of n covariates (x_1, x_2, \dots, x_n) and t refers to the survival time, the coefficients $(\beta_1, \beta_2, \dots, \beta_n)$ measure the impact (i.e., the effect size) of covariates. h_0 describes the baseline hazard. It corresponds to the value of the hazard if all the x_i is equal to zero.

The C-index is calculated as the proportion of consistent outcomes among all possible high-low risk sample pairs (26), which takes values ranging from 0.5 to 1, where 0.5 and 1 represent the worst and best prediction ability, respectively.

Identification of Prognostic Gene Pair Signature

For a gene pair, gene A and gene B, let E_A and E_B represent the expression measurements of gene A and gene B, respectively. And all samples were classified into two groups based on the pattern of REO ($E_A > E_B$ or $E_A < E_B$) in each sample. Using the gene expression of training dataset, we identified a gene pair signature. The detailed process is as follows:

1. The genes whose expression levels are significantly associated with patients' prognoses are identified by using the univariate Cox regression model.
2. Then, for each of the gene pairs, formed from every two of the genes selected above, all samples in the training dataset are divided into two groups according to the REO pattern of the gene pair in each of the samples. If the OS of the two groups of samples are significantly different (univariate Cox regression model), then the REO patterns of the gene pair are considered to be significantly associated with OS, denoted as the pre-selected candidate signature gene pairs. The p -values are adjusted using the Benjamini–Hochberg (BH) procedure (27).
3. From the pre-selected candidate signature gene pairs, the gene pair with the highest C-index value is chosen as the seed signature, and the candidate signature gene pairs are added to the signature one at a time, according to their C-index values ranked in descending order, until the addition of any gene pair does not improve the C-index.
4. A forward selection procedure is performed to search an optimal subset of the candidate signature gene pairs that achieves the highest C-index based on the pre-defined classification rule: a patient is classified into the high-risk group if at least half of the gene pairs of this patient vote for high risk; otherwise, the low-risk group.
5. Finally, the subset of gene pairs with the highest C-index was chosen as the final prognostic gene pair signature.

Analysis of Transcriptional, Epigenomic, and Genomic Data

The Significance Analysis of Microarrays (SAM) method was used to select differentially expressed genes (DEGs) between two groups of samples from BLCA57 and BLCA114 datasets, respectively. Because the RankCompV2 algorithm for detecting DEGs is insensitive to batch effects (28), we use this algorithm to analyze samples from the dataset BLCA158 spread over 15 batches.

The Wilcoxon rank-sum test was used to identify differentially methylated genes (DMGs) between the low- and high-risk samples. Fisher's exact test was used to detect copy number alternations or gene mutations which have significantly different frequencies between two prognostic groups.

Concordance Scores

A concordance score was defined to evaluate the consistency between two lists of DEGs separately detected from two independent datasets. For the two lists of DEGs, if there are k overlapping DEGs, of which s genes showed the same dysregulation directions (overexpressed or underexpressed), the

concordance score was calculated as the ratio, s/k . If k genes were both hypermethylated (or hypomethylated) and differentially expressed, of which s genes were correspondingly underexpressed (or overexpressed), the concordance score was calculated as s/k . This score was used to evaluate the concordance between DEGs and DMGs. The probability of observing a concordance score of s/k by chance is evaluated by the cumulative binomial distribution model (29):

$$P = 1 - \sum_{i=0}^{s-1} \binom{k}{i} (P_0)^i (1 - P_0)^{k-i}$$

where p_0 (here, $p_0 = 0.5$) is the probability of a gene having the concordant relationship between the two gene lists by random chance.

Pathway Enrichment Analysis

GoFunction algorithm was performed to select GO biological pathways that significantly enriched with DEGs. The BH procedure (27) was used to calculate the False Discovery Rate (FDR). All statistical analyses were done by using the R software package version 3.1.3.

RESULTS

Development of the REOs-Based Prognostic Signature

Using the 158 BLCA samples with no drug treatment from the BLCA158 dataset measured by Illumina-Hiseq, we identified 76 genes whose expression values were significantly correlated with the OS of the 158 BLCA patients (univariate Cox proportional-hazards regression model, $FDR < 5\%$). For all gene pairs combined by the 76 genes, 521 OS-associated gene pairs were selected according to their REOs in each sample (univariate Cox regression model, $FDR < 1\%$). For each of the 521 pre-selected candidate signature gene pairs, we classified each of the 158 samples from the BLCA158 dataset into the low- or high-risk group according to the REO of the gene pair in this sample and calculated the C-index value to evaluate the prognostic prediction power of the gene pair (see section Materials and Methods). Then, with the forward selection procedure described in section Materials and Methods, we identified a set of 12 gene pairs that reached the highest C-index value (C-index = 0.80) with the majority voting rule (see section Materials and Methods). The 12 gene pairs, denoted as 12-GPS (**Table 2**), classified patients into the high-risk group if at least six gene pairs vote for high risk; otherwise, the low-risk group (see section Materials and Methods).

Accordingly, patients in the BLCA158 dataset were classified into the low-risk (103 samples) and high-risk (55 samples) groups, which the OS of the former group is significantly better than the latter group (HR = 14.19, 95% CI: 7.66~26.32, $p = 1.18 \times 10^{-25}$, **Figure 1A**). In the 121 samples from the BLCA158 dataset, the DFS of low-risk group (94 samples) is significantly better than the high-risk group (27 samples) (HR = 6.72, 95% CI: 3.62~12.48, $p = 6.99 \times 10^{-12}$, **Figure 1B**). Especially, the 12-GPS classified 51 and 10 patients of the 61 stage I-II patients into the low- and high-risk groups, respectively, and the OS

TABLE 2 | The 12-GPS prognostic signature.

Pair	Gene A	Gene B	Beta	C-index
Pair1	CRAMP1L	BAIAP2	1.21	0.68
Pair2	TALDO1	EMP1	1.40	0.65
Pair3	ALG3	TSPAN7	1.22	0.64
Pair4	ANGEL2	DDX10	1.31	0.64
Pair5	PSMG1	SLC7A11	2.38	0.61
Pair6	CRAMP1L	DDX10	2.07	0.59
Pair7	PRPF4B	UTP6	1.78	0.58
Pair8	TIA1	PSMD11	1.28	0.57
Pair9	ALYREF	ALG3	1.18	0.54
Pair10	PDPK1	ANGPT1	1.3	0.54
Pair11	PSD4	POPDC3	1.67	0.53
Pair12	KMT2B	PSMG1	1.97	0.53

Beta calculated by the univariate Cox regression model represents the risk coefficient of within-sample REO of gene pair (A, B), where $\text{Beta} > 0$ indicates that $E_A < E_B$ is a risk factor.

of the former group is significantly increased compared to the latter group (HR = 46.7, 95% CI: 5.55~393, $p = 1.94 \times 10^{-09}$, **Figure 2A**). The signature could also classify the 58 stage III samples into the low-risk group (35 samples) and the high-risk group (23 samples) with a significantly different OS (HR = 10.30, 95% CI: 4.08~26.49, $p = 8.34 \times 10^{-09}$, **Figure 2B**). Besides, the 12-GPS remained significantly associated with patients' OS after adjusting for AJCC stage, grade, gender and age (multivariate Cox proportional-hazards regression model, HR=11.26, 95% CI: 5.68~22.32, $p=3.89 \times 10^{-12}$, **Table 3**).

Validation of the Signature

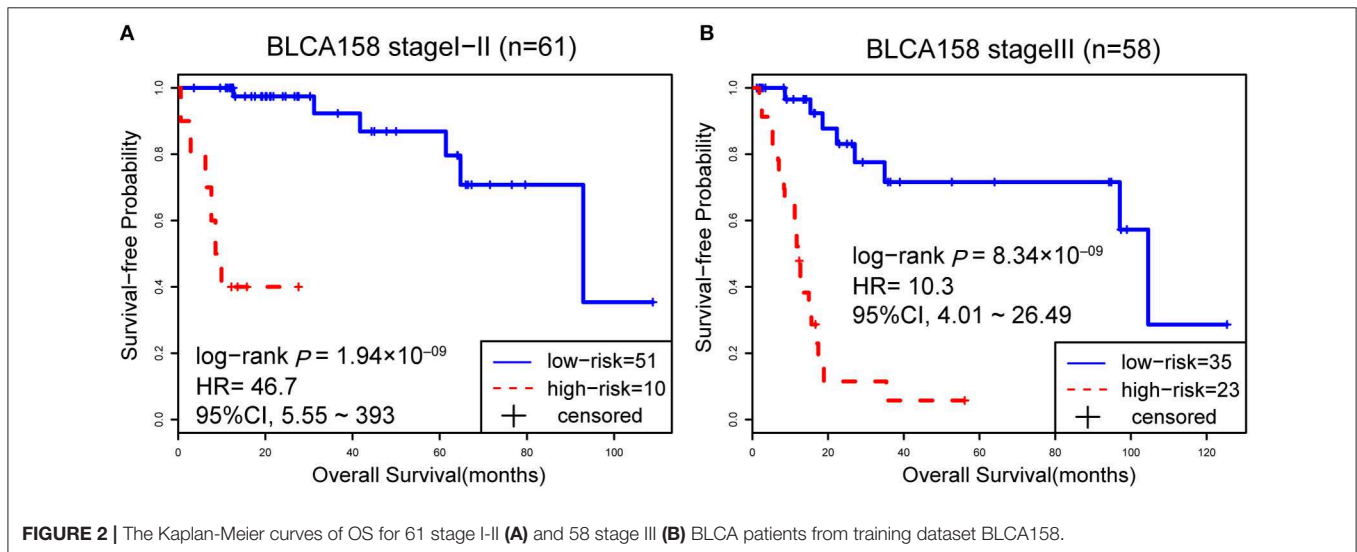
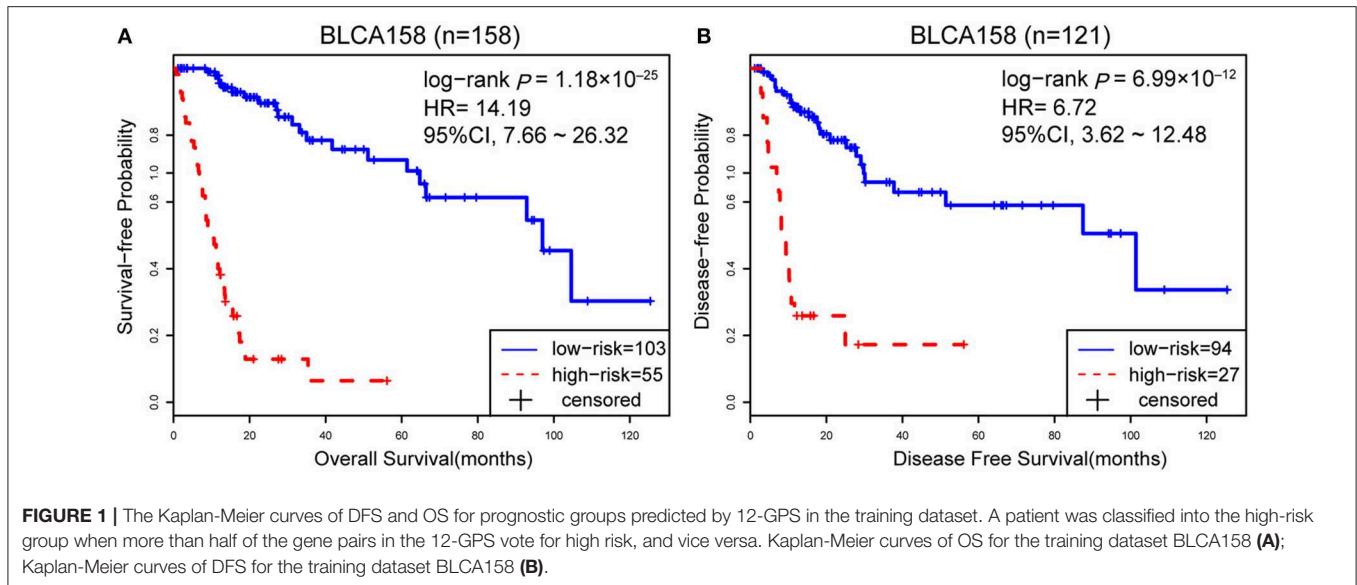
The 114 samples from the BLCA114 dataset were used as the first validation dataset. The signature classified 54 and 60 patients into the low- and high-risk groups, respectively. The low-risk group had a prolonged OS as compared to the high-risk group (HR = 3.59, 95% CI: 1.34~9.62, $p = 6.61 \times 10^{-03}$, **Figure 3A**). The second validation dataset was composed of 57 samples of patients with surgical alone from the BLCA57 dataset. Using the 12-GPS, 12 and 45 samples were classified into the low- and high-risk groups with a marginally significantly different OS but extremely high HR (HR = 2.69×10^{08} , 95% CI: 0~Inf, $p = 0.06$, **Figure 3B**) mainly due to the small sample size. For the 57 samples with DFS data in this dataset, the low-risk group (12 samples) have a significantly better DFS than the high-risk group (45 samples) (HR = 2.75×10^{08} , 95% CI: 0~Inf, $p = 0.05$, **Figure 3C**). Especially, the 12 patients of the low-risk group had no death or recurrence during the follow-up period.

However, due to the lack of DFS information for the patients from the dataset BLCA114, we were only able to analyze the OS for the union of the dataset BLCA57 and BLCA114. For the combined data of the two validation datasets, the 12-GPS classified 57 stage I-II patients into the low-risk group and 70 stage I-II patients into the high-risk group, with significantly different OS (HR = 3.41, 95% CI: 1.15~10.11, $p = 0.02$, **Figure 4A**). And the OS of the low-risk patients of stage III

stratified by the 12-GPS remained better than that of the high-risk patients although the difference was not significant which would be due to the limited number ($n = 30$) of samples in this dataset (HR = 3.50, 95% CI: 0.045~27.43, $p = 0.20$, **Figure 4B**). The biomarker proposed in this study is to predict the recurrence risk of patients only treated with curative surgery. And, this signature was developed by assuming that the existence of occult micro-metastases would cause stage I to III patients to have a relapse after curative surgery. That is, 12-GPS might not entirely be suitable for stage IV bladder cancer. In the combined data of BLCA158 and BLCA57, 48 stage IV patients were available and among them, 4 patients had distant metastases. As expected, we found that the developed signature was able to classify the four distantly metastatic patients into the high-risk group. The stratification of the other 48 stage IV patients was also statistically significant (HR = 6.53, 95% CI: 2.20~19.42, $p = 1.32 \times 10^{-04}$, **Figure S1**). Although 12-GPS was not initially designed for stage IV bladder cancer, it could predict the prognosis of these even more aggressive patients to some extent. Besides, within the high-risk group, the samples were further classified into the highest-risk group (Count ($E_A < E_B$) ≥ 9 , **Table 2**) if at least 9 gene pairs of 12-GPS vote for high risk, which had the poorer prognosis than those classified into the middle-high-risk group ($9 > \text{Count} (E_A < E_B) \geq 6$, **Table 2**). Survival analysis for the integrated data of the three datasets showed that, as the threshold of gene pair number for supporting high risk increased to 9, the predicted highest-risk patients had even worse OS than the middle-high- and low-risk patients (HR = 1.86, 95% CI: 1.50~2.31, $p = 1.56 \times 10^{-10}$, **Figure 5A**). Similarly, in the combination of the datasets BLCA158 and BLCA57, the patients with the highest-risk had shorter DFS than the middle-high- and low-risk patients (HR = 1.52, 95% CI: 1.17~1.97, $p = 1.80 \times 10^{-04}$, **Figure 5B**). In the dataset BLCA57, the grade (reflecting the degree of tumor invasion) of a patient was evaluated as grade 1, grade 2 or grade 3, whereas in the dataset BLCA114 the grade of a patient was evaluated as high grade or low grade. Here, according to the Malignancy Grading of Bladder Carcinoma: Old and New Systems in NCCN Guidelines Version 1.2017 Bladder Cancer, in the unified dataset of BLCA 57 and BLCA114, we defined grade 1 as the low grade, and grade 3 as the high grade, excluding grade 2 samples from the analysis to avoid confusion. After adjusting for the AJCC stage, grade, age and gender, the 12-GPS remained significantly associated with patients' OS (multivariate Cox proportional-hazards regression model, HR = 3.39, 95% CI: 1.01~11.39, $p = 0.049$, **Table 3**) in the integrated data.

Distinct Transcriptional and Epigenomic Characteristics of the Prognostic Subtypes

Between the stage IV (N+, M+ or T4b) samples and stage I-III (T1-T4aN0M0) samples from the BLCA114 and BLCA57 datasets, no DEG was identified (SAM, FDR < 1%). Between the stage IV samples and stage I-III samples from the BLCA158 dataset, 73 DEGs were identified using the RankCompV2 algorithm (FDR < 1%). Thus, the signals of differential expression between the primary tumors with metastasis and non-metastasis based on AJCC stage alone are weak and irreproducible in



independent datasets, which is due to the high false positive and false negative reports of the commonly used imaging techniques and pathological examinations for tumor metastasis (6).

In contrast, we identified 2,171 and 2,356 DEGs (SAM, FDR < 1%) between the low- and high-risk groups from the samples of the BLCA114 and BLCA57 datasets, respectively. Similarly, 922 DEGs were identified from the dataset BLCA158 using the RankCompV2 algorithm (FDR < 1%). The consistency of DEGs between every two datasets is higher than 99.63% (binomial distribution test, $p < 2.2 \times 10^{-16}$, **Table 4**), suggesting significantly consistent transcriptional differences between the low- and high-risk groups.

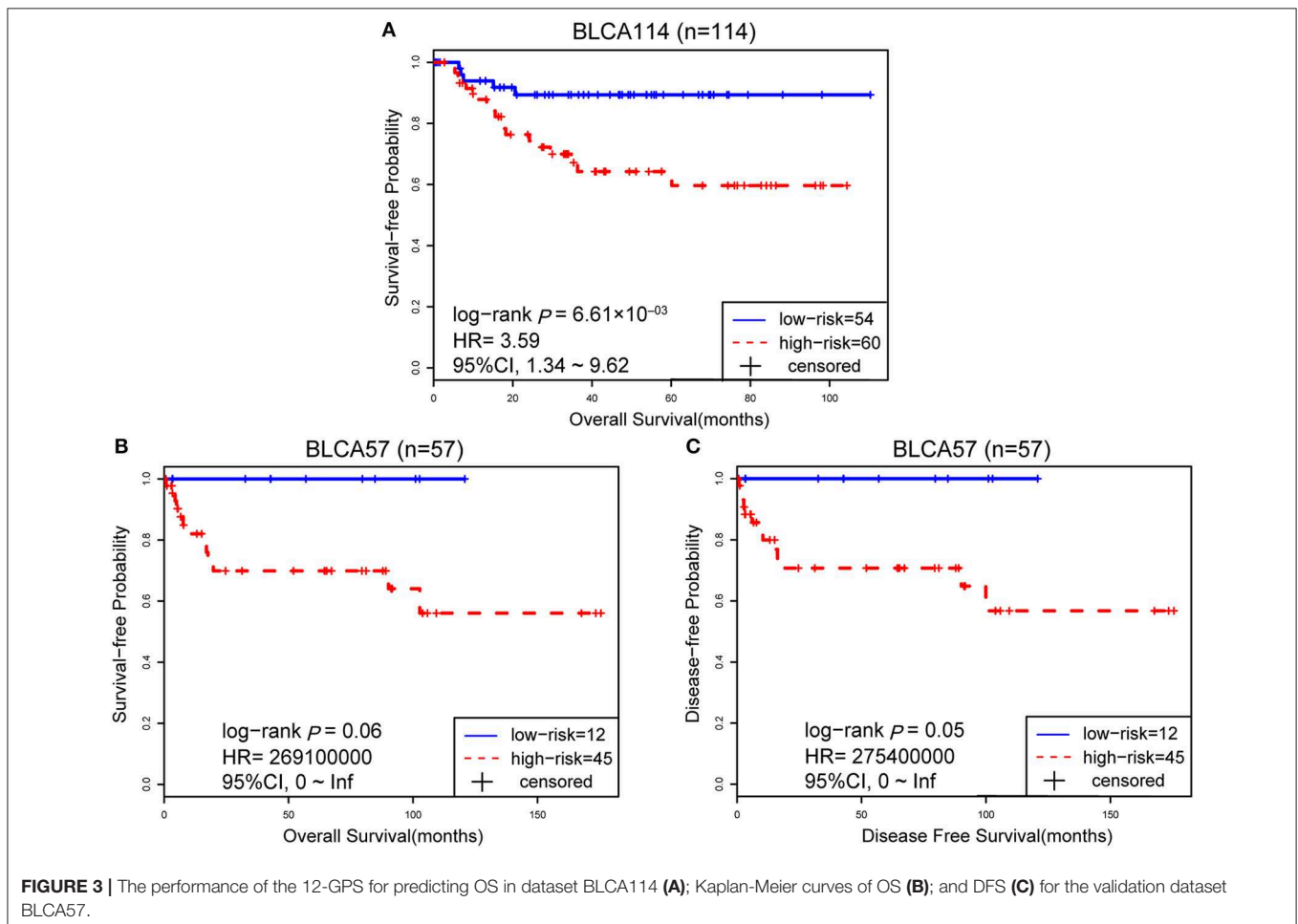
By pooling the DEGs detected in the three datasets, we obtained 4,061 unique DEGs when excluding three DEGs with contradictory dysregulation directions between two datasets. Functional enrichment analysis showed that the 4,061 DEGs

were significantly enriched in some pathways typically related to tumor metastasis such as “cell adhesion” (30), “cell migration” (30, 31), “cell differentiation” (32), and “cell division” (33) (FDR < 1%, hypergeometric distribution model, **Table S1**).

According to the reclassified samples in BLCA158 dataset by 12-GPS, 605 hypermethylated and 736 hypomethylated genes were identified by comparing the methylation profiles of BLCA158 dataset between the two prognostic groups (Wilcoxon rank-sum test, FDR < 1%), respectively. Two hundred and three genes out of the 605 hypermethylated genes were also overlapped with 4,061 DEGs, among which 94.09% were concordantly underexpressed in the high-risk group compared with the low-risk group, which was unlikely to happen by chance (binomial distribution test, $p < 2.2 \times 10^{-16}$). Similarly, we found that there are 284 overlaps between the 736 hypomethylated genes and the DEGs, among which 94.72% were concordantly overexpressed in

TABLE 3 | Univariate and multivariate Cox regression analyses for the 12-GPS signature.

Variables	Univariate model		Multivariate model	
	HR (95% CI)	<i>p</i>	HR (95% CI)	<i>p</i>
BLCA158				
Predictive signature (high vs. low)	14.09 (7.60 ~ 26.14)	2.00×10^{-16}	11.89 (6.13 ~ 23.09)	2.60×10^{-13}
Stage (I vs. II vs. III vs. IV)	2.40 (1.73 ~ 3.32)	3.00×10^{-08}	1.66 (1.14 ~ 2.41)	0.0081
Gender (male vs. female)	0.73 (0.44 ~ 1.23)	0.20	0.76 (0.46 ~ 1.28)	0.31
Age	1.05 (1.02 ~ 1.07)	8.00×10^{-04}	1.05 (1.01 ~ 1.08)	0.0042
Grade (high vs. low)	2.91 (0.40 ~ 21.17)	0.30	0.19 (0.02 ~ 1.63)	0.13
The unified data of BLCA54 and BLCA114				
Predictive signature (high vs. low)	4.31 (1.32 ~ 14.11)	0.0085	3.39 (1.01 ~ 11.39)	0.049
Stage (I vs. II vs. III vs. IV)	2.12 (1.46 ~ 3.07)	4.92×10^{-05}	2.02 (1.35 ~ 3.01)	5.79×10^{-04}
Age	0.98 (0.94 ~ 1.02)	0.36	0.99 (0.97 ~ 1.01)	0.20
Gender (male vs. female)	1.44 (0.60 ~ 3.49)	0.41	1.32 (0.54 ~ 3.21)	0.54
Grade (high vs. low)	2.00 (0.27 ~ 14.67)	0.49	0.61 (0.07 ~ 5.00)	0.64



the high-risk group, which was also highly unlikely to happen by chance (binomial distribution test, $p < 2.2 \times 10^{-16}$). Notably, using Fisher's exact test with FDR < 1%, we were unable to detect genes which had significantly different copy number alternation frequencies or mutation frequencies between the two prognostic groups.

DISCUSSION

In this study, we developed a qualitative transcriptional signature based on the within-sample REOs of 12-GPS for predicting the DFS and OS of stage I-III BLCA patients after surgical resection. As mentioned in the Introduction, this REOs-based

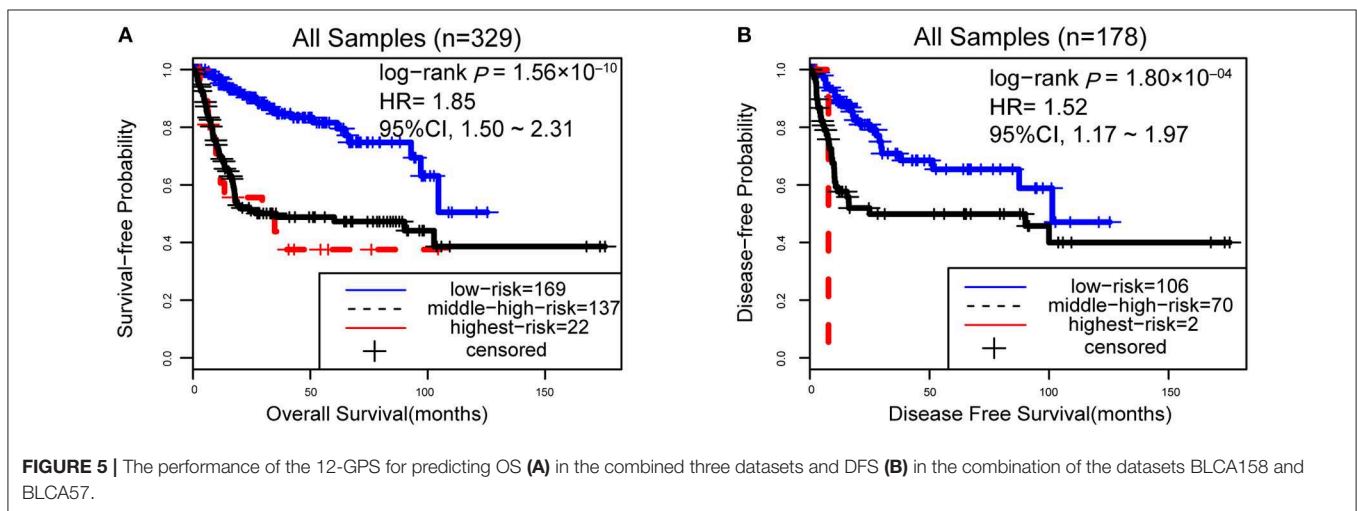
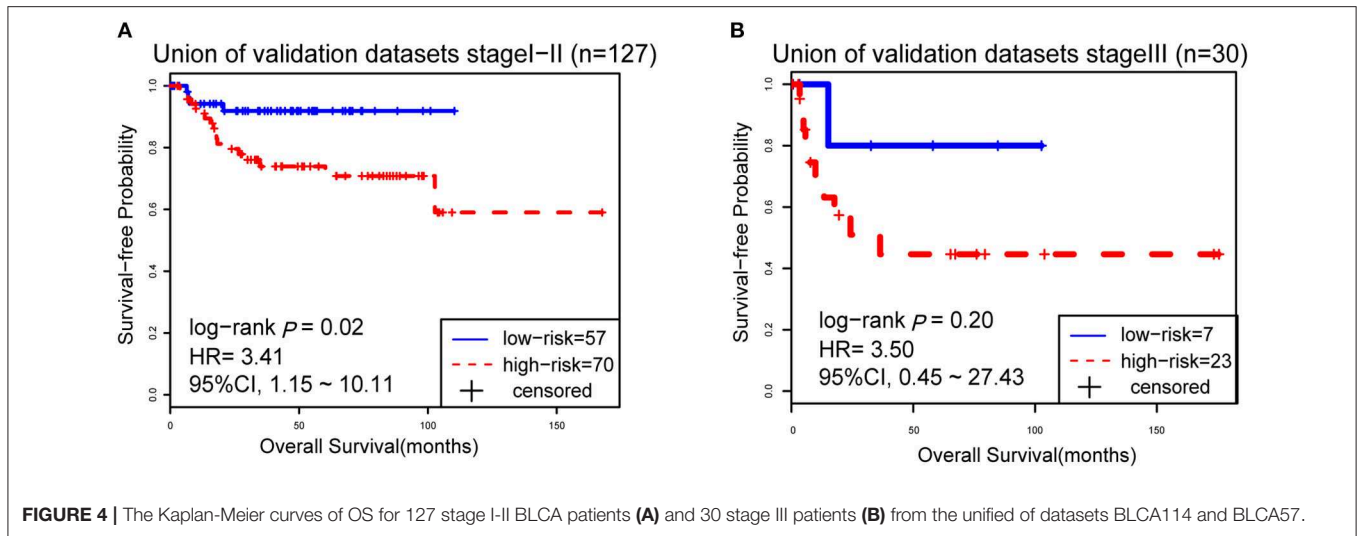


TABLE 4 | Concordance scores between DEGs detected from different datasets.

Comparisons	Overlap	Consistence	Concordance scores	p-value
BLCA158 vs. BLCA114	475	475	100%	$<2.20 \times 10^{-16}$
BLCA158 vs. BLCA57	350	350	100%	$<2.20 \times 10^{-16}$
BLCA114 vs. BLCA57	800	797	99.63%	$<2.20 \times 10^{-16}$

qualitative signature is highly robust against experimental batch effects, data normalization, varied proportions of the tumor epithelial cell in tumor tissues sampled from different tumor locations of the same patient (18), partial RNA degradation during specimen storage and preparation (17) and amplification bias for minimum specimens even with about 15–25 cancer cells (19). Therefore, it can be robustly applied at the individual level to samples measured by different laboratories. Remarkably, if treated with curative surgery only, the patients with low recurrence risk, classified by the signature, could not benefit from adjuvant therapy. After excluding these patients with adjuvant

therapy-irrelevant low recurrence risk and collecting enough samples of patients treated with a particular neoadjuvant or adjuvant therapy, we can establish a predictive signature for the patients of high-risk group to identify which patients would response to a particular neoadjuvant or adjuvant therapy (34, 35).

For the establishment of predictive signatures for stage I-III bladder cancer patients who would be treated with curative surgery, we need to firstly identify the patients with adjuvant therapy-irrelevant low recurrence risk after surgery and exclude them from the analysis for predictive signatures because they would confound the identification of predictive signatures of response to a specific adjuvant therapy (34, 36). Therefore, the identification of prognostic signature for stage I-III bladder cancer patients treated with curative surgery only should be a basis for the establishment of predictive signatures of response to a specific adjuvant therapy. Next, we will develop a predictive signature to identify which patients could response to or benefit from a specific adjuvant therapy including chemotherapy or radiotherapy after collecting enough data of patients treated with a specific adjuvant therapy.

This qualitative signature can also help to determine the metastasis status under clinical settings based on the hypothesis that occult micro-metastases contribute to high recurrence risk of patients after surgical resection (5). Besides, we found several signature genes have important biological meaning associated with the carcinogenesis. For instance, TALDO1, as a nearly ubiquitous enzyme, has been linked to the progression of bladder cancer (37). Another signature gene, ALG3, has a higher expression in samples of breast cancer with advanced stages than those with early stages (38) and the decreased expression of PSD4 (EFA6B) was associated with poor prognosis of patients with breast cancer (39). Through the regulation of cellular invasiveness and migration, ALYREF is linked to local lymph node metastasis in human oral squamous cell carcinoma (40). Further, with the aid of 12-GPS, the tumor samples in the low- and high-risk groups showed a significantly consistent transcriptional differences characteristics related to tumor metastasis in independent datasets. Functional enrichment analysis showed that these transcriptional differences were significantly enriched in some classical metastasis-associated pathways, including “cell adhesion” (30), “cell migration” (30, 31), and “cell differentiation” (32). Further, we found that metastasis-associated DNA methylation alterations were significantly concordant with transcriptional differences observed between the low- and high-risk groups, indicating that DNA methylation alternations of the CpG loci in these genes play important roles in promoting cancer metastasis.

There are more than 60% of non-muscle invasive patients would recur and ~50% muscle invasive patients would develop metastases after curative surgery (2, 41, 42). It would be reasonable to assume that a major common factor, the existence of occult micro-metastases, leading to recurrence after curative surgery for the patients with the stage of localized/locally advanced, which would influence the choices of subsequent treatment for the bladder cancer patients. So, although the management is usually different between non-muscle invasive and muscle invasive cancer, we pooled them together into our analysis, while the later (muscle invasive cancer) would have higher probabilities of harboring occult micro-metastases and thus provide additional information for the discovery and validation of the signature for predicting the prognosis or recurrence risk (potential micro-metastasis) of patients after curative surgery (43).

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* (2015) 65:87–108. doi: 10.3322/caac.21262
2. Sylvester RJ, van der Meijden AP, Oosterlinck W, Witjes JA, Bouffoux C, Denis L, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur Urol.* (2006) 49:466–5. doi: 10.1016/j.eururo.2005.12.031
3. Lughezzani G, Sun M, Shariat SF, Budaus L, Thuret R, Jeldres C, et al. A population-based competing-risks analysis of the survival of patients treated with radical cystectomy for bladder cancer. *Cancer.* (2011) 117:103–9. doi: 10.1002/cncr.25345

In summary, given that the subtle quantitative measurements of gene expression are unreliable (15), the apparent shortcoming of this qualitative signature that it might lose some subtle quantitative information of gene expression is actually a unique merit of robustness. Therefore, it is worthy to be further verified in clinical practice.

CONCLUSIONS

The REO-based 12-GPS prognostic signature is a true individual-level qualitative signature, which can robustly identify the stage I-III BLCA patients with potential occult micro-metastases and be an auxiliary tool for clinicians to determine whether patients should receive adjuvant therapy after surgical resection. Compared with the low-risk samples, the high-risk samples identified by 12-GPS have distinct transcriptional and epigenetic features characterizing BLCA metastasis.

AUTHOR CONTRIBUTIONS

ZG and YL: conception and design. HZ, YG, HC, XL, and QG: development of methodology. YL and JH: acquisition of data. ZG, YL, YG, and HZ: analysis and interpretation of data. YL, ZG, H-ML, and XW: writing, review, and/or revision of the manuscript. ZG: study supervision. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant numbers: 81572935 and 21534008), the Joint Scientific and Technology Innovation Fund of Fujian Province (Grant number: 2016Y9044).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2019.00629/full#supplementary-material>

Figure S1 | The Kaplan-Meier curves of OS for 47 stage IV BLCA patients in the unified data of BLCA158 and BLCA57 datasets.

Table S1 | Functional analysis for 4061 DEGs (FDR<1%, hypergeometric distribution model).

4. Mitra AP, Quinn DI, Dorff TB, Skinner EC, Schuckman AK, Miranda G, et al. Factors influencing post-recurrence survival in bladder cancer following radical cystectomy. *BJU Int.* (2012) 109:846–54. doi: 10.1111/j.1464-410X.2011.10455.x
5. Jensen JB, Hoyer S, Jensen KM. Incidence of occult lymph-node metastasis missed by standard pathological examination in patients with bladder cancer undergoing radical cystectomy. *Scand J Urol Nephrol.* (2011) 45:419–24. doi: 10.3109/00365599.2011.599336
6. Li M, Hong G, Cheng J, Li J, Cai H, Li X, et al. Identifying reproducible molecular biomarkers for gastric cancer metastasis with the aid of recurrence information. *Sci Rep.* (2016) 6:24869. doi: 10.1038/srep24869
7. Salminen AP, Jambor I, Syvanen KT, Bostrom PJ. Update on novel imaging techniques for the detection of lymph node

- metastases in bladder cancer. *Minerva Urol Nefrol.* (2016) 68:138–49.
8. Zargar H, Zargar-Shoshtari K, Dundee P, Black PC. Predicting occult lymph node-positive disease at the time of radical cystectomy: a systematic review. *Minerva Urol Nefrol.* (2016) 68:112–24.
 9. Herr HW. Extent of surgery and pathology evaluation has an impact on bladder cancer outcomes after radical cystectomy. *Urology.* (2003) 61:105–8. doi: 10.1016/S0090-4295(02)02116-7
 10. Jeong P, Ha YS, Cho IC, Yun SJ, Yoo ES, Kim IY, et al. Three-gene signature predicts disease progression of non-muscle invasive bladder cancer. *Oncol Lett.* (2011) 2:679–84. doi: 10.3892/ol.2011.309
 11. Kim WJ, Kim SK, Jeong P, Yun SJ, Cho IC, Kim IY, et al. A four-gene signature predicts disease progression in muscle invasive bladder cancer. *Mol Med.* (2011) 17:478–85. doi: 10.2119/molmed.2010.00274
 12. Mitra AP, Lam LL, Ghadessi M, Erho N, Vergara IA, Alshalalfa M, et al. Discovery and validation of novel expression signature for postcystectomy recurrence in high-risk bladder cancer. *J Natl Cancer Inst.* (2014) 106:dju290. doi: 10.1093/jnci/dju290
 13. Chu J, Li N, Li F. A risk score staging system based on the expression of seven genes predicts the outcome of bladder cancer. *Oncol Lett.* (2018) 16:2091–6. doi: 10.3892/ol.2018.8904
 14. Qi L, Chen L, Li Y, Qin Y, Pan R, Zhao W, et al. Critical limitations of prognostic signatures based on risk scores summarized from gene expression levels: a case study for resected stage I non-small-cell lung cancer. *Brief Bioinform.* (2016) 17:233–42. doi: 10.1093/bib/bv064
 15. Guan Q, Yan H, Chen Y, Zheng B, Cai H, He J, et al. Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal cancer. *BMC Genomics.* (2018) 19:99. doi: 10.1186/s12864-018-4446-y
 16. Xu H, Guo X, Sun Q, Zhang M, Qi L, Li Y, et al. The influence of cancer tissue sampling on the identification of cancer characteristics. *Sci Rep.* (2015) 5:15474. doi: 10.1038/srep15474
 17. Chen R, Guan Q, Cheng J, He J, Liu H, Cai H, et al. Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. *Oncotarget.* (2017) 8:6652–62. doi: 10.18632/oncotarget.14257
 18. Cheng J, Guo Y, Gao Q, Li H, Yan H, Li M, et al. Circumvent the uncertainty in the applications of transcriptional signatures to tumor tissues sampled from different tumor sites. *Oncotarget.* (2017) 8:30265–75. doi: 10.18632/oncotarget.15754
 19. Liu H, Li Y, He J, Guan Q, Chen R, Yan H, et al. Robust transcriptional signatures for low-input RNA samples based on relative expression orderings. *BMC Genomics.* (2017) 18:913. doi: 10.1186/s12864-017-4280-7
 20. Patil P, Bachant-Winner PO, Haibe-Kains B, Leek JT. Test set bias affects reproducibility of gene signatures. *Bioinformatics.* (2015) 31:2318–23. doi: 10.1093/bioinformatics/btv157
 21. Zhang L, Hao C, Shen X, Hong G, Li H, Zhou X, et al. Rank-based predictors for response and prognosis of neoadjuvant taxane-anthracycline-based chemotherapy in breast cancer. *Breast Cancer Res Treat.* (2013) 139:361–9. doi: 10.1007/s10549-013-2566-2
 22. Ao L, Song X, Li X, Tong M, Guo Y, Li J, et al. An individualized prognostic signature and multiomics distinction for early stage hepatocellular carcinoma patients with surgical resection. *Oncotarget.* (2016) 7:24097–110. doi: 10.18632/oncotarget.8212
 23. Qi L, Li T, Shi G, Wang J, Li X, Zhang S, et al. An individualized gene expression signature for prediction of lung adenocarcinoma metastases. *Mol Oncol.* (2017) 11:1630–45. doi: 10.1002/1878-0261.12137
 24. Ao L, Zhang Z, Guan Q, Guo Y, Guo Y, Zhang J, et al. A qualitative signature for early diagnosis of hepatocellular carcinoma based on relative expression orderings. *Liver Int.* (2018) 38:1812–19. doi: 10.1111/liv.13864
 25. Jones MP, Crowley J. A general class of nonparametric tests for survival analysis. *Biometrics.* (1989) 45:157–70. doi: 10.2307/2532042
 26. Harrell FEJr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* (1996) 15:361–87.
 27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing[J]. *J R Stat Soc.* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
 28. Cai H, Li X, Li J, Liang Q, Zheng W, Guan Q, et al. Identifying differentially expressed genes from cross-site integrated data based on relative expression orderings. *Int J Biol Sci.* (2018) 14:892–900. doi: 10.7150/ijbs.24548
 29. Bahn AK. Application of binomial distribution to medicine: comparison of one sample proportion to an expected proportion (for small samples). Evaluation of a new treatment evaluation of a risk factor. *J Am Med Womens Assoc.* (1969) 24:957–66.
 30. Sun Y, Cheng MK, Griffiths TR, Mellon JK, Kai B, Kriajevska M, et al. Inhibition of STAT signalling in bladder cancer by diindolylmethane: relevance to cell adhesion, migration and proliferation. *Curr Cancer Drug Targets.* (2013) 13:57–68. doi: 10.2174/156800913804486610
 31. Li Y, Ishiguro H, Kawahara T, Kashiwagi E, Izumi K, Miyamoto H. Loss of GATA3 in bladder cancer promotes cell migration and invasion. *Cancer Biol Ther.* (2014) 15:428–35. doi: 10.4161/cbt.27631
 32. Tornetta FJ. Clinical studies with the new antiemetic, metoclopramide. *Anesth Analg.* (1969) 48:198–204. doi: 10.1213/0000539-196903000-00008
 33. Preston-Martin S, Pike MC, Ross RK, Jones PA, Henderson BE. Increased cell division as a cause of human cancer. *Cancer Res.* (1990) 50:7415–21.
 34. Cai H, Li X, Li J, Ao L, Yan H, Tong M, et al. Tamoxifen therapy benefit predictive signature coupled with prognostic signature of post-operative recurrent risk for early stage ER+ breast cancer. *Oncotarget.* (2015) 6:44593–608. doi: 10.18632/oncotarget.6260
 35. Guo Y, Jiang W, Ao L, Song K, Chen H, Guan Q, et al. A qualitative signature for predicting pathological response to neoadjuvant chemoradiation in locally advanced rectal cancers. *Radiother Oncol.* (2018) 129:149–53. doi: 10.1016/j.radonc.2018.01.010
 36. Song K, Guo Y, Wang X, Cai H, Zheng W, Li N, et al. Transcriptional signatures for coupled predictions of stage II and III colorectal cancer metastasis and fluorouracil-based adjuvant chemotherapy benefit. *FASEB J.* (2019) 33:151–62. doi: 10.1096/fj.201800222RRR
 37. Silvers CR, Miyamoto H, Messing EM, Netto GJ, Lee YF. Characterization of urinary extracellular vesicle proteins in muscle-invasive bladder cancer. *Oncotarget.* (2017) 8:91199–208. doi: 10.18632/oncotarget.20043
 38. Yang Y, Zhou Y, Xiong X, Huang M, Ying X, Wang M. ALG3 is activated by heat shock factor 2 and promotes breast cancer growth. *Med Sci Monit.* (2018) 24:3479–87. doi: 10.12659/MSM.907461
 39. Zangari J, Partisani M, Bertucci F, Milanini J, Bidaut G, Berruyer-Pouyet C, et al. EFA6B antagonizes breast cancer. *Cancer Res.* (2014) 74:5493–506. doi: 10.1158/0008-5472.CAN-14-0298
 40. Saito Y, Kasamatsu A, Yamamoto A, Shimizu T, Yokoe H, Sakamoto Y, et al. ALY as a potential contributor to metastasis in human oral squamous cell carcinoma. *J Cancer Res Clin Oncol.* (2013) 139:585–94. doi: 10.1007/s00432-012-1361-5
 41. Millan-Rodriguez F, Chechile-Toniolo G, Salvador-Bayarri J, Palou J, Algaba F, Vicente-Rodriguez J. Primary superficial bladder cancer risk groups according to progression, mortality and recurrence. *J Urol.* (2000) 164:680–4. doi: 10.1016/S0022-5347(05)67280-1
 42. Stein JP, Lieskovsky G, Cote R, Grossen S, Feng AC, Boyd S, et al. Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. *J Clin Oncol.* (2001) 19:666–75. doi: 10.1200/JCO.2001.19.3.666
 43. Zhao W, Chen B, Guo X, Wang R, Chang Z, Dong Y, et al. A rank-based transcriptional signature for predicting relapse risk of stage II colorectal cancer identified with proper data sources. *Oncotarget.* (2016) 7:19060–71. doi: 10.18632/oncotarget.7956
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Zhang, Guo, Cai, Li, He, Lai, Guan, Wang and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.