Research Paper

Development and validation of an intra-tumor heterogeneity-related signature to predict prognosis of bladder cancer: a study based on single-cell RNA-seq

Ranran Zhou[1,2], Jingjing Liang[3], Qi Chen[1,2], Hu Tian[1,2], Cheng Yang[1,2], Cundong Liu[1,2]

[1]Department of Urology, The Third Affiliated Hospital of Southern Medical University, Guangzhou, China
[2]The Third School of Clinical Medicine, Southern Medical University, Guangzhou, China
[3]Department of Cardiology, Shunde Hospital of Southern Medical University, Foshan, China

Correspondence to: Cundong Liu; **email:** cundongliu@163.com, https://orcid.org/0000-0002-2098-1139
Keywords: intra-tumor heterogeneity, bladder cancer, scRNA-seq, prognosis, risk model
Received: May 9, 2021 **Accepted:** July 15, 2021 **Published:** August 2, 2021

ABSTRACT

Intra-tumor heterogeneity (ITH) was a potential mechanism of progression and drug resistance in bladder cancer (BCa). However, the understanding of ITH in BCa remains insufficient. Single-cell RNA sequencing (scRNA-seq) profiles of 2075 cells were analyzed, and 2940 cell markers were screened. The ITH of 396 cases was evaluated, and 96 ITH-related genes were identified. Based on the gene-pair strategy, 96 genes were cyclically paired, and an 8-gene-pair model was successfully established to evaluate the overall survival of BCa through Lasso and multivariate Cox regressions. The risk model showed high predictive value in the training dataset (*n* = 396, *p* = 0) and external validation datasets (*n* = 165, *p* = 2.497e-02; *n* = 224, *p* = 3.423e-02). The model was also valuable for the prediction of clinical treatment outcomes. Totally, a prognostic model based on scRNA-seq and ITH was successfully constructed and validated in large cohorts, providing novel clues for ITH study of BCa.

INTRODUCTION

Bladder cancer (BCa), as the twelfth most common malignancies around the world, brings a tremendous social burden [1]. The 5-year survival rate of muscle-invasive bladder cancer (MIBC), one of the main subtypes of BCa, is dismal: 5%–30% [2]. Nowadays, some novel therapeutic methods, such as cisplatin-based neoadjuvant chemotherapy and immune checkpoint inhibitors (ICIs), have been proposed, making considerable strides in BCa treatment [3]. However, many BCa patients could not benefit from the current therapeutic regimens [4, 5]. Therefore, reliable prediction of prognosis was urgently demanded, which played an important role in guiding clinical treatment.

With the proliferation of tumor cells, the genomic characteristics of progeny cells are different from that of their parents, inducing the alternation of drug susceptibility, invasiveness, migration, and growth, which is known as intra-tumor heterogeneity (ITH) [6]. ITH is closely correlated with immunotherapy response because the neoantigen on tumor cells with high ITH is diluted, and the concentration is not enough to cause the antitumor immunity [7]. ITH is also capable of predicting the prognosis of patients with malignancies [8]. The underlying mechanisms of ITH include telomere damage, DNA mismatch repair deficiency, microsatellite instability (MSI), and epigenetic changes [9], but the understanding of ITH is far from enough for the moment.

Various methods, like fluid biopsy, gene sequencing, and multi-regional biopsy, have been developed to estimate the ITH of BCa patients. Sing-cell RNA sequencing (scRNA-seq) has attracted more and more

attention due to its high resolution [10]. For instance, Maynard et al. disclosed the dynamic changes of lung tumor cells in patients who received target therapy via scRNA-seq technology [11], showing single-cell sequencing was a powerful tool for ITH research.

In the present study, we quantified the ITH of 396 cases with BCa from The Cancer Genome Atlas (TCGA). Then scRNA-seq data was collected from Gene Expression Omnibus (GEO), where we also downloaded the external validation datasets. A gene-pair strategy was implemented to improve the robustness of the established model [12]. The correlation between risk signature and clinical treatment outcomes was also explored. Our research constructed a promising tool to predict the clinical outcomes and provided some novel biomarkers, deepening the understanding of ITH in BCa.

RESULTS

ITH estimation of 396 cases with BCa

The workflow chart of this study is shown in Figure 1A. We utilized the mutant-allele tumor heterogeneity (MATH) algorithm to evaluate ITH of BCa cases from TCGA (Supplementary Table 1). Accordingly, the ITH of each individual was quantified as MATH value, and high MATH represented increased ITH of malignant tumors [13]. It was found that the BCa cases with high MATH suffered a poorer survival rate (p = 4.146e-02, Figure 1B) and lower sensitivity to chemotherapeutic agents (Figure 1C). The predictive potential of MATH to immunotherapy effectiveness was also evaluated. MATH was positively correlated with Tumor Mutational Burden (TMB) via Spearman correlation (r = 0.15, p = 0.0017, Figure 1D). With ESTIMATE algorithm, the proportion of immune and stromal components of the tumor microenvironment (TME) was calculated [14]. The ratios of immune and stromal components in TME in the low-MATH group were significantly higher than those in the high-MATH group by Wilcoxon test (p < 0.001, Figure 1E). The expression level of routine immune checkpoint genes, including *PD1*, *LAG-3*, *GAL-9*, *CTLA-4*, *TIM-3*, and *TIGIT*, were negatively correlated with MATH (p < 0.05, Figure 1F).

Profiling of scRNA-seq data

The transcriptome sequencing of 2075 cells isolated from a patient with primary BCa was obtained from the GEO website [15]. Figure 2A displayed the detected gene numbers, sequencing count, and the percent of mitochondrial genes of each cell. With the sequencing depth increased, the percent of mitochondrial genes (r = –0.64) and detected gene count (r = 0.92) were also significantly changed (Figure 2B). The Top10 genes with the most significant variation across 2075 cells included *TPSB2*, *TPSAB1*, *IGFBP7*, *S100A2*, *CD74*, *HLA-DRA*, *MALAT1*, *HLA-DRB1*, *PLA2G2A*, and *FN1* (Figure 2C). The principal component analysis (PCA) was conducted to classify the cells preliminarily (Figure 2D). The p-value of each principal component (PC) was illustrated in Figure 2E, and the correlated genes of Top4 PC were shown in Supplementary Figure 1. To get a more precise clustering of cell samples, t-Distributed Stochastic Neighbor Embedding (t-SNE) was then implemented, and 2075 cell samples were divided into 14 different clusters (Supplementary Table 2, Figure 2F). A total of 2940 markers genes were screened with |logFC| > 0.5 and adjusted p < 0.05 (Supplementary Table 3), and accordingly, the cell types were annotated (Figure 2F). Figure 2G indicated the trajectory analysis of 14 cell clusters, which re-validated the annotation of cell types.

ITH-related genes screening and heterogeneity-related score (HRS) construction

We calculated the Spearman correlation coefficients between MATH values and transcriptome expression levels of 2940 cell markers, and 96 genes were screened (r > 0.25, p < 0.05, Supplementary Table 4, Figure 3A). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional annotation indicated the 96 genes mainly were enriched in the biological process related to the cell cycle (Figure 3B and 3C). After cyclically pairing the screened 96 genes, 1256 gene pairs were identified. Subsequently, 10 ITH-related gene-pairs were extracted with the Lasso-Cox algorithm (Figure 3D and 3E), 8 of which were included in the prognostic model via multivariate Cox regression with stepwise (Supplementary Table 5, Figure 3F). Here, we defined the risk score, which was calculated based on the established risk model, as a heterogeneity-related score (HRS). To help clinicians better understand HRS, a nomogram was drawn (Figure 4A). The expression level of 13 genes, which comprised the 8-gene-pair model, in 14 different cell subgroups were sown in Figure 3G. Figure 3H illustrated the mutational rates of the genes in different tumor pathological stages.

Validation of HRS

The calibration analysis indicated the estimated 3- (Figure 4B) and 5-year (Figure 4C) OS rates from HRS were close to the actual survival rates. The optimal HRS value, which was detected through X-tile software [16], was equal to 2.16, according to which each case was labeled with low- or high-risk (Supplementary Table 6). Compared with the patients with low-risk, the individuals with high-risk suffered significantly poorer

prognosis (p = 0, Figure 4D). With HRS increasing, more deaths were observed (Figure 4G and 4J). Besides, HRS could be used to evaluate the survival rate in almost all subgroups by stratification survival analysis (Supplementary Figure 2). The external validation was also conducted in GSE13507 and GSE328094 datasets, which included 165 and 224 cases, respectively. The details of clinicopathological features of the datasets enrolled were s shown in Table 1. Based on the optimal cut-off of 2.16, the patients from these two researches were divided into low- and high-risk subgroups (Supplementary Tables 7 and 8). Kaplan-Meier survival plots displayed HRS could effectively predict the survival rates in GSE13507 (p = 2.497e-02, Figure 4E)

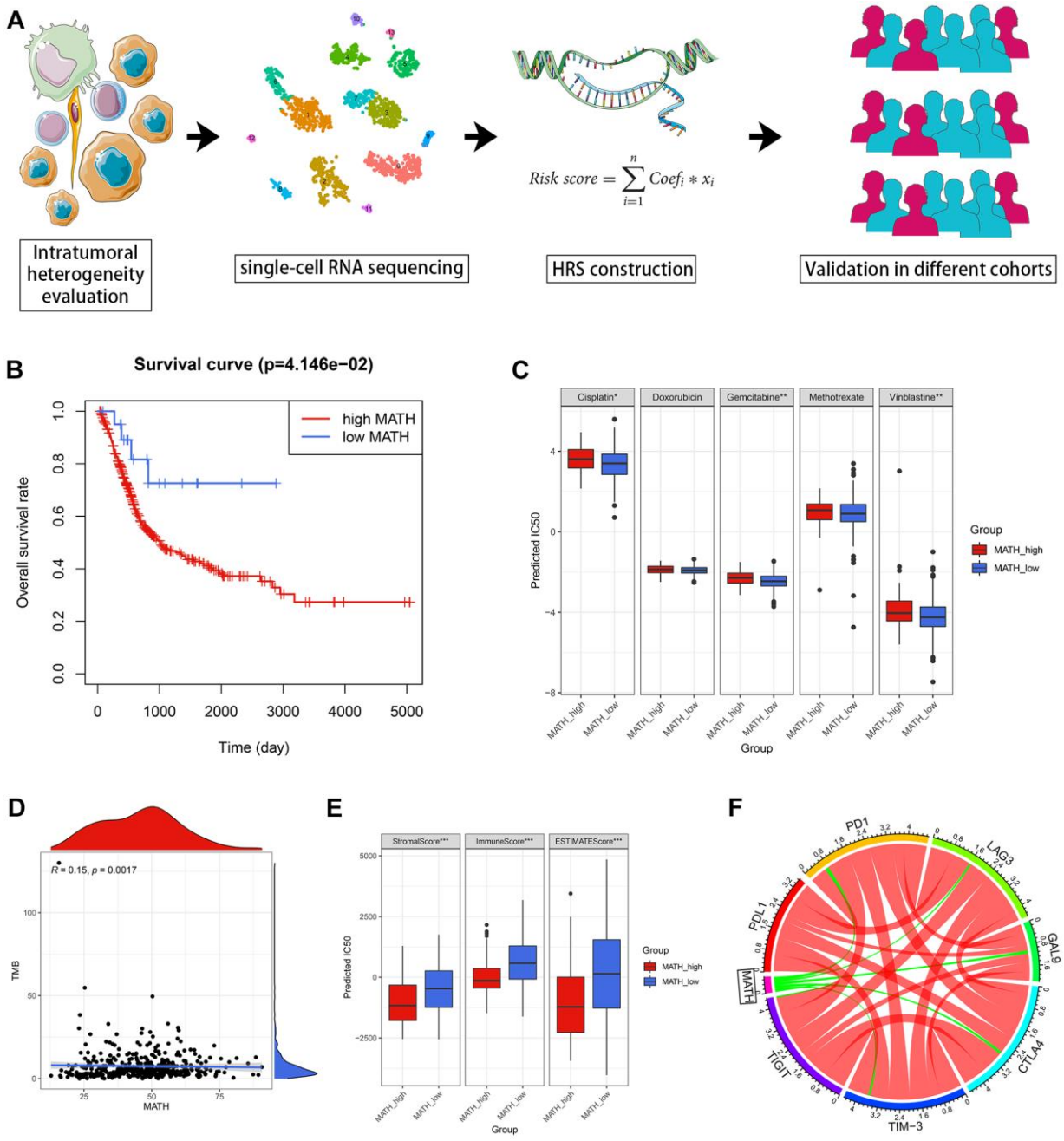


Figure 1. Evaluation of ITH with MATH in BCa. (**A**) The workflow of this study. (**B**) The patients with high MATH values suffered an unfavorable prognosis. (**C**) The estimated MATH values acted as a potential predictor for chemosensitivity with Wilcoxon signer-rank test. (**D**) MATH values were positively correlated with TMB. (**E**) The cases in the high-MATH group had significantly lower immune and stromal components in TME. (**F**) Spearman correlation analysis indicated MATH values were negatively correlated with routine immune checkpoint genes, including *PD1, LAG3, GAL9, CTLA4, TIM-3,* and *TIGIT*. The red lines and green lines represented positive correlation and negative correlation, respectively. The boldness of the lines was positively associated with the strength of the correlation. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

and GSE32894 (p = 3.423-02, Figure 4F) cohorts. The distribution of HRS and survival status of GSE13507 (Figure 4H and 4K) and GSE32894 (Figure 4I and 4L) were also analyzed and illustrated. Wilcoxon signed-rank test conformed that the 13 genes involved in HRS were mostly differentially expressed between adjacent normal and BCa tissues (Supplementary Figure 3), and most of them could predict OS with significant efficacy (Supplementary Figures 4 and 5, Supplementary Tables 9–11). Besides, we also compared the mRNA expression level of the 13 genes between normal urothelium cell line SV-HUC-1 and BCa cell line T24 via Real-time quantitative PCR (RT-qPCR, Supplementary Figure 6), and the sequences of primers utilized are shown in Supplementary Table 12.

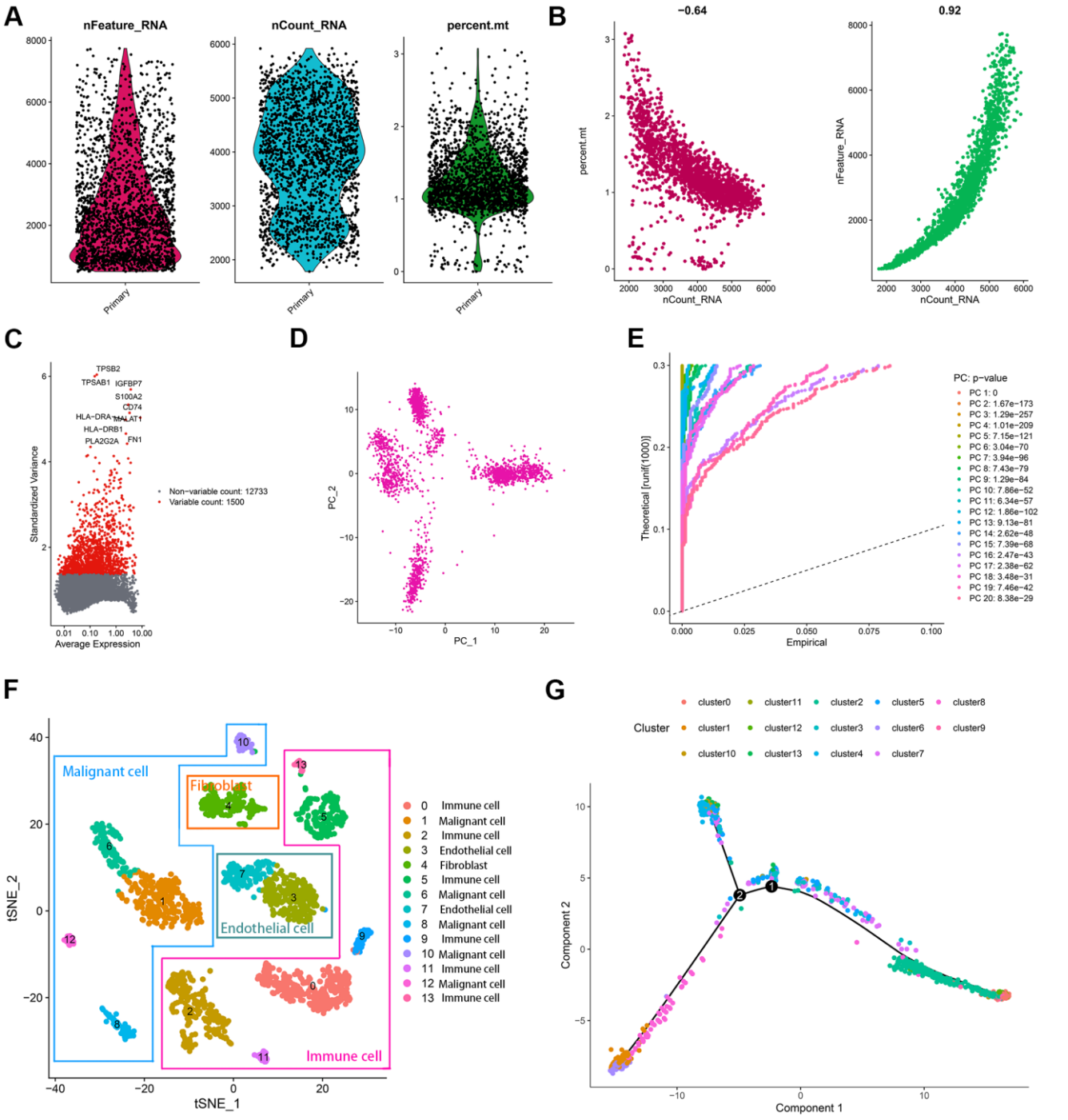


Figure 2. Characterization of scRNA-seq from 2075 cells. (**A**) Quality control plots of cell samples. (**B**) The sequencing depth was negatively correlated with the proportion of mitochondrial genes and positively associated with detected gene numbers. (**C**) 1500 variable genes across cell samples were identified. (**D**, **E**) PCA was conducted to reduce the dimension of data sets. (**F**) Cell samples were classified into 14 clusters with the t-SNE algorithm. (**G**) The trajectory analysis of 14 cell clusters.

Clinical evaluation by HRS

Compared with other clinicopathological features, like age, gender, pathological stages, pathological T stages, M stages, and pathological N stages, HRS was an independent prognostic factor no matter in the univariate Cox (Figure 5A) and multivariate Cox (Figure 5B) analysis. The areas under curves (AUCs) of HRS were also higher than those clinical parameters in 1-year (Figure 5C), 3-year (Figure 5D) and 5-year

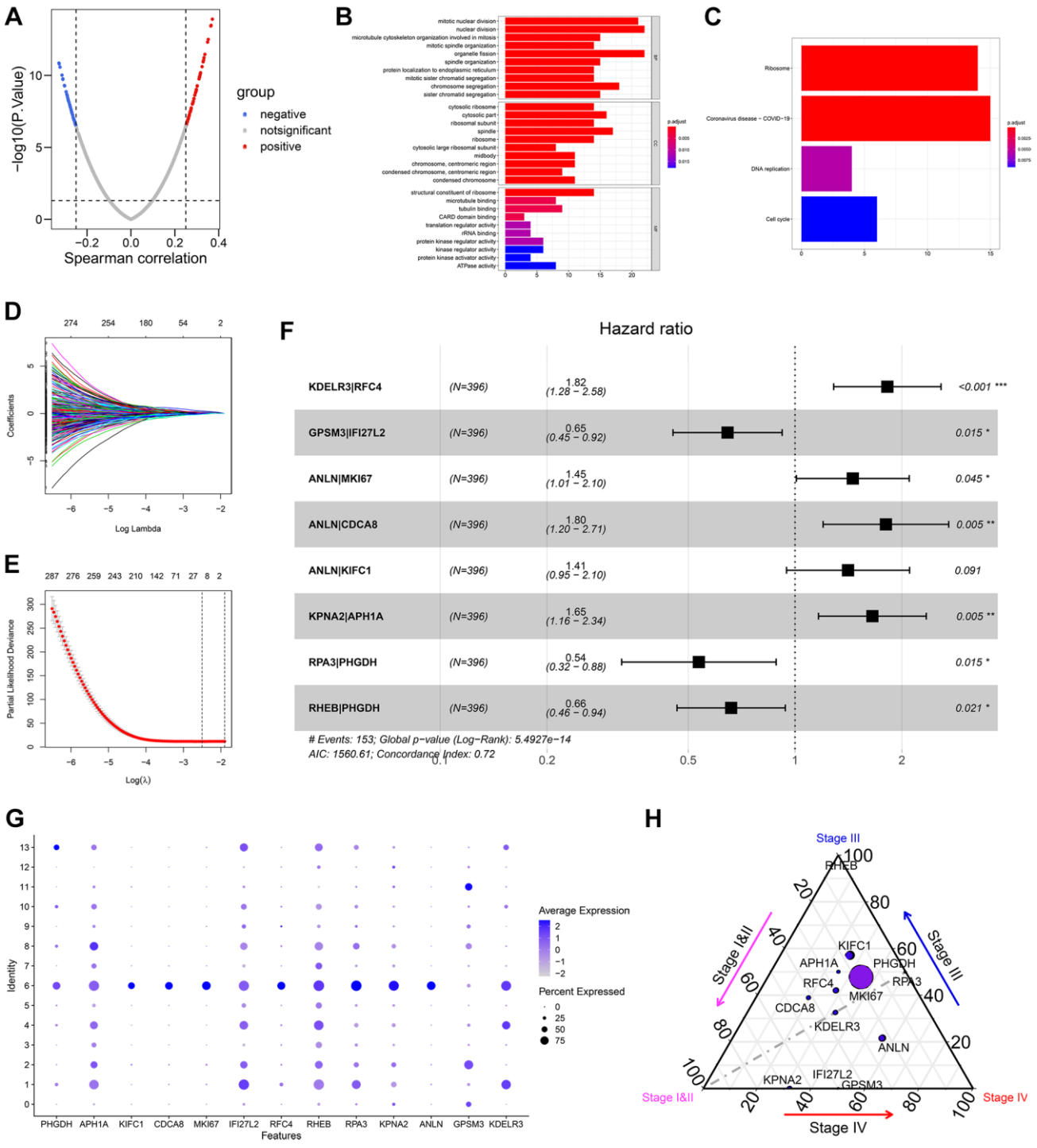


Figure 3. Establishment of HRS. (**A**) 96 of 2075 cell marker genes were significantly correlated with MATH. (**B**, **C**) GO functional annotation (**B**) and KEGG pathway enrichment (**C**) of the 96 genes. (**D**, **E**) 10 crucial gene-pairs correlated with OS were identified via Lasso-Cox regression. (**F**) 8 of 10 gene pairs were included in the prognostic model by multivariate Cox analysis with stepwise. (**G**) The expression values of 13 genes comprising the 8-gene-pair signature in 14 cell subpopulations. (**H**) The mutational rates of the 13 genes in different stages in BCa. The size of the bubble represented the mutational rates in all samples.

(Figure 5E) receiver operating characteristic (ROC) curves. The strip curve indicated HRS was significantly associated with pathological tumor stages via the Chi-square test ($p < 0.05$, Figure 6A). Wilcoxon test displayed that the HRS was also correlated with other clinicopathological traits (Figure 6B–6G), which were widely regarded as risk factors for BCa.

Correlation between HRS and clinical treatment

HRS was significantly associated with MATH values through Wilcoxon signed-rank test ($p = 0.0037$, Figure 7A) and Spearman correlation analysis (r = 0.17, $p = 0.00095$, Figure 7B). We also found that the high-HRS group had a relatively lower *GAL-9* expression level

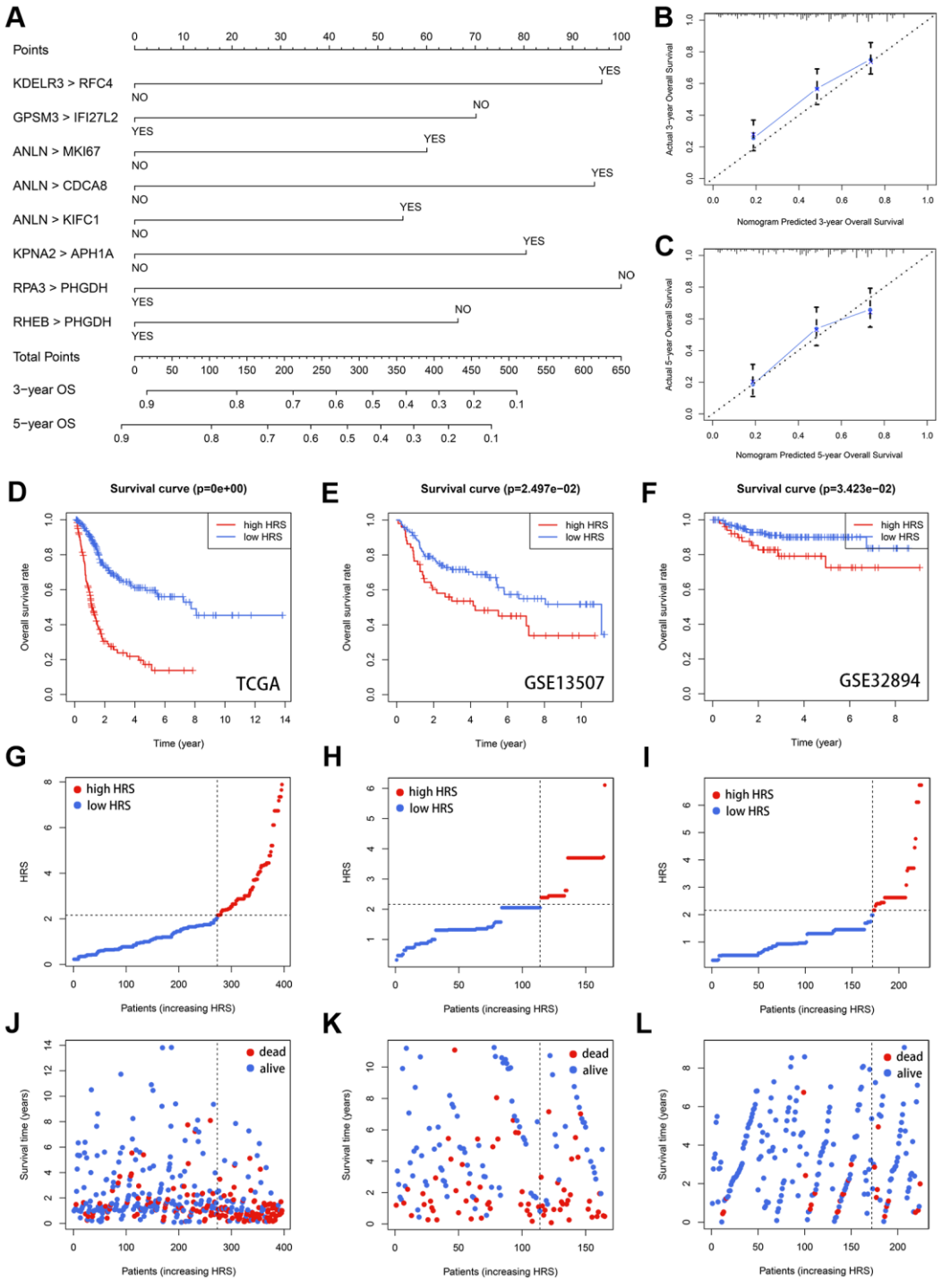


Figure 4. Validation of HRS. (**A**) A nomogram was plotted to visualize the HRS. (**B**, **C**) The calibration curves for 3- (**B**) and 5-year (**C**) OS prediction. (**D**–**F**) Kaplan-Meier survival analysis of HRS in TCGA (**D**), GSE13507 (**E**) and GSE32894 (**F**) cohorts. (**G**–**I**) The distribution of HRS in TCGA (**G**), GSE13507 (**H**) and GSE32894 (**I**). (**J**–**L**) The distribution of survival status in TCGA (**J**), GSE13507 (**K**) and GSE32894 (**L**).

Table 1. The baseline information of 785 cases enrolled in the present study.

Parameters	TCGA (*n* = 396)	GSE13507 (*n* = 165)	GSE32894 (*n* = 224)
Survival status			
Alive	243 (61.3%)	96 (58.1%)	199 (88.8%)
Dead	153 (38.6%)	69 (41.8%)	25 (11.1%)
Follow-up (day)	778.19 ± 814.38	1451.45 ± 1127.70	1196.98 ± 767.38
Age	67.84 ± 10.53	65.18 ± 11.93	69.43 ± 11.28
Gender			
Female	104 (26.2%)	30 (18.1%)	61 (27.2%)
Male	292 (73.7%)	135 (81.8%)	163 (72.7%)
Pathological Stage			
I	2 (0.5%)	–	–
II	124 (31.3%)	–	–
III	138 (34.8%)	–	–
IV	130 (32.8%)	–	–
Unknown	2 (0.5%)	–	–
pT stage			
T0	1 (0.2%)	0 (0.0%)	0 (0.0%)
Ta	0 (0.0%)	23 (13.9%)	110 (49.1%)
T1	3 (0.7%)	81 (49.0%)	63 (28.1%)
T2	113 (28.5%)	31 (18.7%)	43 (19.1%)
T3	190 (47.9%)	19 (11.5%)	7 (3.1%)
T4	57 (14.3%)	11 (6.6%)	1 (0.4%)
Unknown	32 (8.0%)	0 (0.0%)	0 (0.0%)
M stage			
M0	189 (47.7%)	158 (95.7%)	–
M1	10 (2.5%)	7 (4.2%)	–
Unknown	197 (49.7%)	0 (0.0%)	–
pN stage			
N0	229 (57.8%)	149 (90.3%)	27 (12.0%)
N1	44 (11.1%)	8 (4.8%)	3 (1.3%)
N2	75 (18.9%)	6 (3.6%)	10 (4.4%)
N3	7 (1.7%)	1 (0.6%)	0 (0.0%)
Unknown	41 (10.3%)	1 (0.6%)	184 (82.1%)
Risk stratification			
High	123 (31.0%)	51 (30.9%)	52 (23.2%)
Low	273 (68.9%)	114 (69.0%)	172 (76.7%)
HRS	1.90 ± 1.58	1.96 ± 1.03	1.52 ± 1.20

Abbreviations: TCGA: The Cancer Genome Atlas; HRS: heterogeneity-related score.

than the low-HRS group via the Wilcoxon test ($p < 0.001$, Figure 7C). With the CIBERSORT algorithm [17], we estimated the infiltration proportion of 22 immune cells in TME and found some cell types, including $CD8^+$ T cells, Tregs, M0 macrophages, resting mast cells, and neutrophils, were closely

associated with HRS by Wilcoxon test ($p < 0.05$, Figure 7D). Most of the HRS-related immune cells have been reported to influence immunotherapeutic outcomes [18–20]. It was also found that the evaluated chemotherapy sensitivity to cisplatin, methotrexate, and vinblastine was different in the low-HRS and high-HRS subgroup by means of the Wilcoxon test ($p < 0.05$, Figure 7E). In all, HRS had the potential to serve as a biomarker for the effectiveness of clinical treatment, containing immunotherapy and chemotherapy, in BCa.

HRS-related biological pathways detection

Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) were both performed to ensure the predictive accuracy of the pathway enrichment results. The analysis results of GSVA were shown in Figure 8A and Supplementary Table 13, and GSEA results were supplemented in Supplementary Tables 14 and 15. Among the related biological process, G2M checkpoint, mitotic spindle, mTORC1 signaling, and epithelial-mesenchymal transition (EMT) were observed to be positively correlated with HRS (Figure 8B and 8C), while DNA repair and oxidative phosphorylation (OXPHOS) had a significantly negative association (Figure 8D and 8E).

The candidate compounds targeting HRS

To analyze the potential HRS-related compounds, we uploaded the differentially-expressed genes (DEGs) extracted from high-HRS and low-HRS patients and illustrated in Figure 9A and Supplementary Table 16 to the CMap database. Seven potential compounds, including cephaeline (Figure 9C), LY-294002 (Figure 9D), lycorine (Figure 9E), naltrexone (Figure 9F), nefopam (Figure 9G), tanespimycin (Figure 9H), and wortmannin (Figure 9I), were the predicted small molecule drugs with the most significance (Figure 9B, Supplementary Table 17).

DISCUSSION

BCa is among the most malignant tumors around the world. The prognosis of BCa is unfavorable, especially for the cases with locally advanced cancers and distant metastasis [21]. Therefore, the accurate prediction of BCa prognosis remains one of the topics we are concerned about most. ITH, a hallmark of malignant cancers, refers to the change of molecular biology and genomic features in tumor cell evolution [22]. It has been widely accepted that ITH acted as one of the fundamental causes of many phenotypes of cancers. Recent research characterized ITH across 38 different tumors, uncovering some latent driving factors [9].

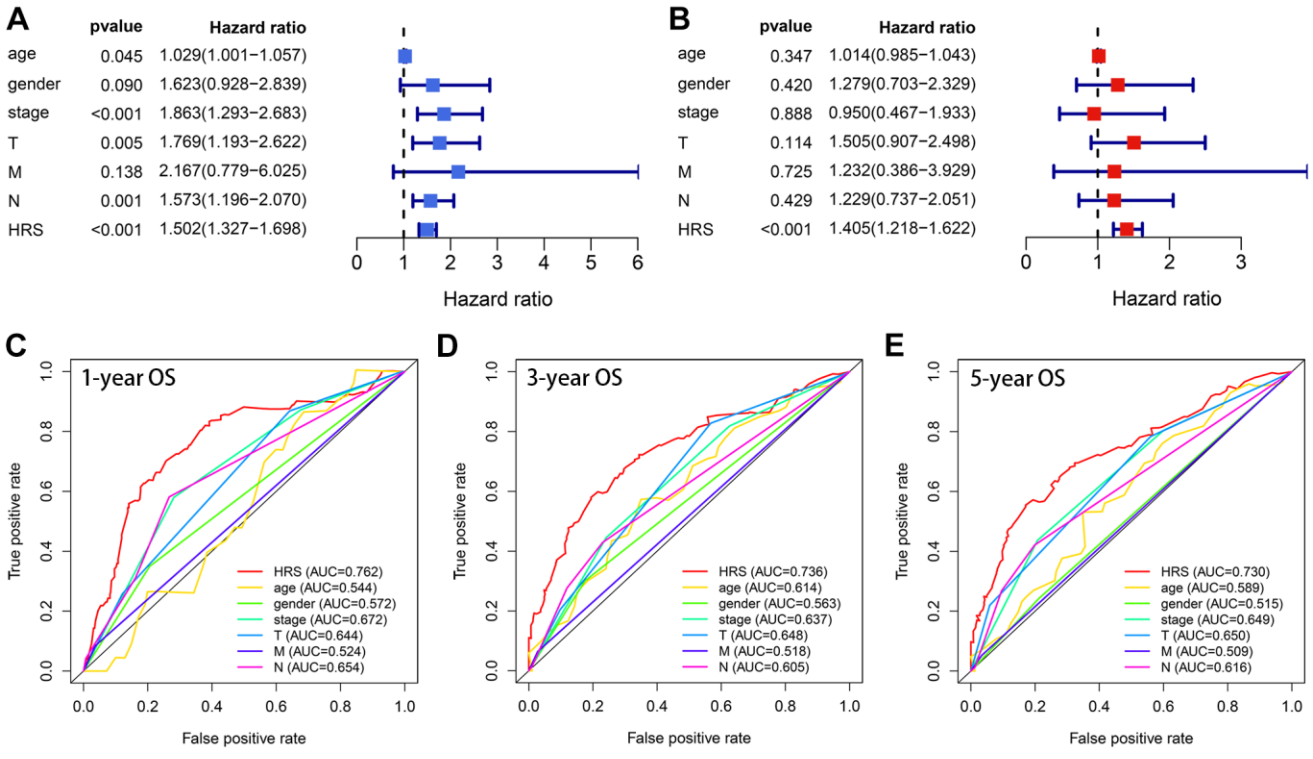


Figure 5. HRS showed superiority in OS prediction compared with the clinicopathological features. (**A**, **B**) HRS was an independent prognosis predictor in univariate (**A**) and multivariate (**B**) analyses. (**C**–**E**) ROC analysis indicated HRS had better ability than the clinical traits in 1- (**C**), 3- (**D**) and 5-year (**E**) OS prediction.

Nevertheless, it is a long and complex process to comprehensively understand ITH, and more studies should be conducted. At present, more and more evolution traits of malignant cells have been revealed based on single-cell level [23], demonstrating single-cell technology was a valuable maneuver to analyze ITH.

Some previous studies contributed to the prognosis prediction of BCa. For example, according to TP53 mutation status, Wu et al. developed and validated a predictive model for OS of BCa [24]. Based on glucose metabolism, an 18-gene signature was successfully built to predict prognosis in urothelial carcinoma [25]. These efforts helped for individualized treatment and the underlying mechanisms exploring. However, the prognostic value of ITH-related signature has not been reported in BCa. Besides, most of the previous studies were based on the DEGs identified from the tissues with different statuses, like tumor and paracarcinoma, or high and low immune infiltration, which might omit important biomarkers because of the ignorance of the

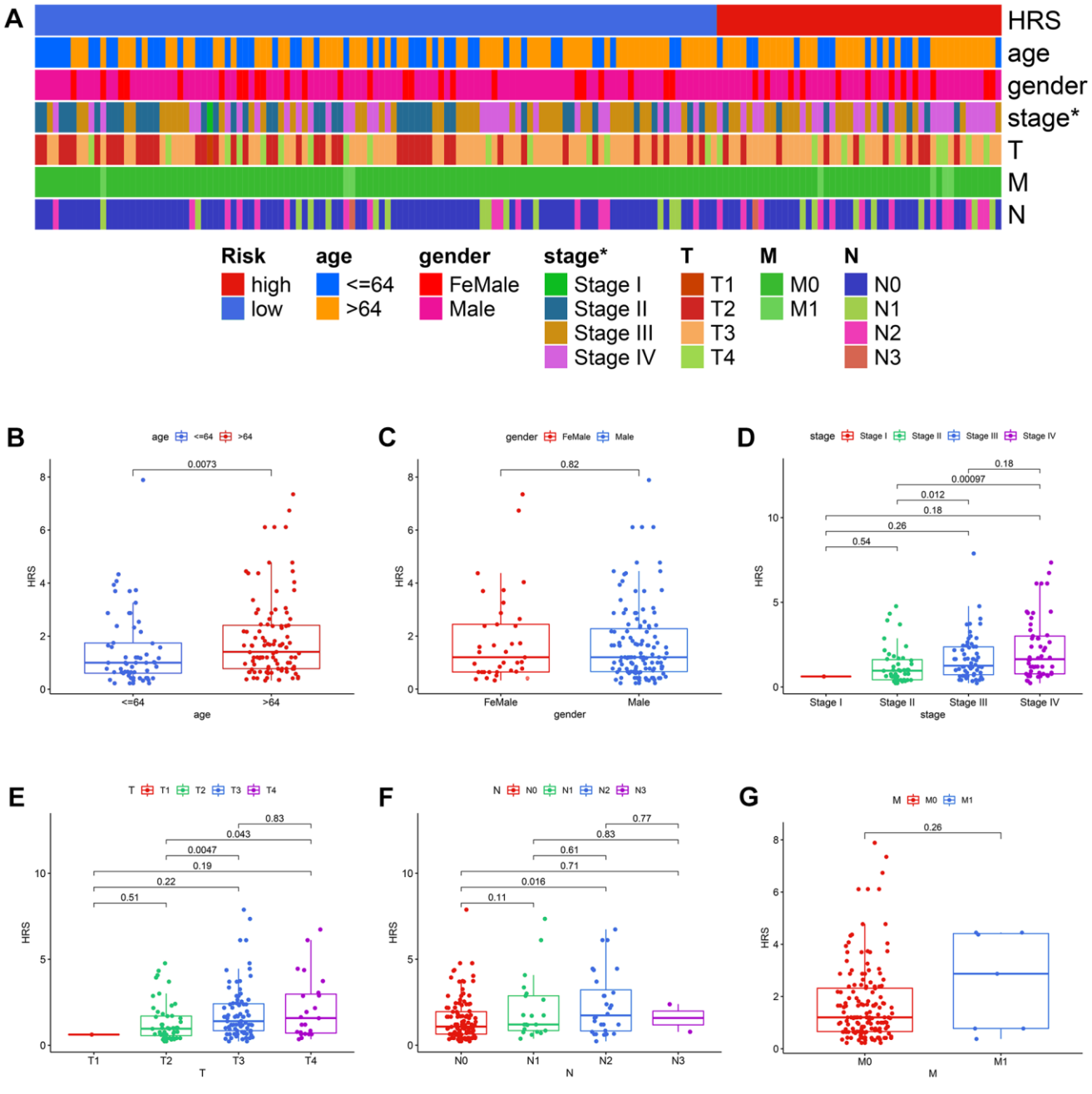


Figure 6. The correlation between HRS and other clinicopathological variables. (**A**) The strip curve displayed HRS was significantly correlated with pathological tumor stages. (**B**–**G**) Correlation analysis of HRS with age (**B**), gender (**C**), pathological stages (**D**), pathological T stages (**E**), pathological N stages (**F**), and M stages (**G**).

differences among various cell subsets. Given the reason mentioned above, some researchers have constructed and successfully validated the scRNA-seq-based prognostic models in cancers [26, 27], but no risk signature on the basis of scRNA-seq has developed in BCa. Therefore, the present study established an ITH-related prognosis signature based on scRNA-seq, which was reasonable and urgently demanded.

This study first evaluated the ITH of 396 BCa cases from the TCGA-BLCA dataset via the MATH algorithm, which was developed based on mutant-allele fractions (MAFs) and validated in head and neck squamous cell carcinoma [8, 13]. The MATH values were positively correlated with ITH, as well as the unfavorable survival rate (Figure 1B). Then scRNA-seq data of 2075 cells obtained from a patient with primary BCa was analyzed, and 14 different cell types were identified. Based on the cell markers of cell clusters, 96 genes were screened as ITH-related genes via co-expression analysis with MATH values. GO and KEGG analysis displayed that the genes mainly were enriched in cell cycle-related pathways. Subsequently, 96 genes were cyclically paired, and a 0-or-1 matrix was built with gene-pair strategy, where there was no need for definite expression level of genes. Lasso and multivariate Cox with stepwise helped establish an 8-gene-pair prognostic model, and accordingly, the heterogeneity-related score (HRS) was calculated for each individual. The cases with high HRS suffered poorer survival in the TCGA-BLCA, GSE13507, and GSE32895 datasets, indicating that HRS might serve as a reliable risk predictor. Besides, the expression value of *GAL-9* and $CD8^+$ T cell infiltration proportion in the high-HRS group was significantly lower than that in the low-HRS populations, showing HRS might be linked to

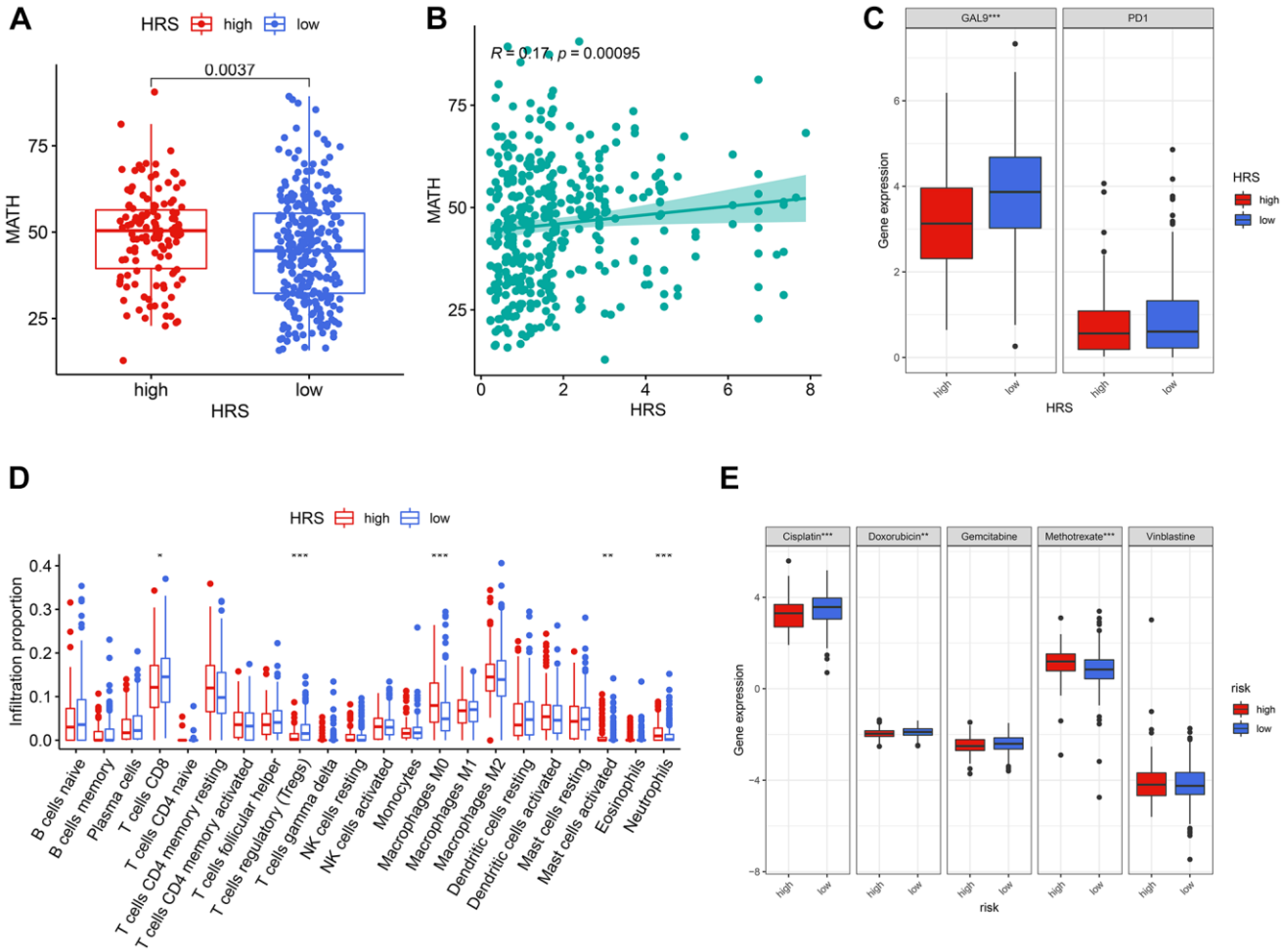


Figure 7. HRS was a potential predictor for clinical treatment of BCa. (**A**, **B**) HRS was significantly correlated with MATH, codetermined by difference (**A**) and correlation (**B**) analysis. (**C**) GAL-9 was differentially expressed between low- and high-HRS groups. (**D**) The cases in the high-HRS group were more likely to be associated with the high infiltration of M0 macrophages, activated mast cells, and neutrophils, whereas they were negatively correlated with the infiltration of $CD8^+$ T cells and Tregs. (**E**) High HRS was linked to a lower IC50 for chemotherapeutics like cisplatin and doxorubicin, whereas it was correlated to a higher IC50 of methotrexate. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

immunotherapy response. It was also found the HSR had the potential to indicate chemotherapy effectiveness, including cisplatin, doxorubicin and methotrexate.

Several biomarkers were firstly reported to be correlated with the development of BCa. *KDELR3*, which was up-regulated in tumor samples and cell lines compared with that in adjacent normal tissues and normal cell lines (Supplementary Figures 3 and 6), served as a predictor for poor prognosis of BCa (Supplementary Figures 4 and 5). It was reported that the loss of *KDELR3* in the genetically engineered mouse

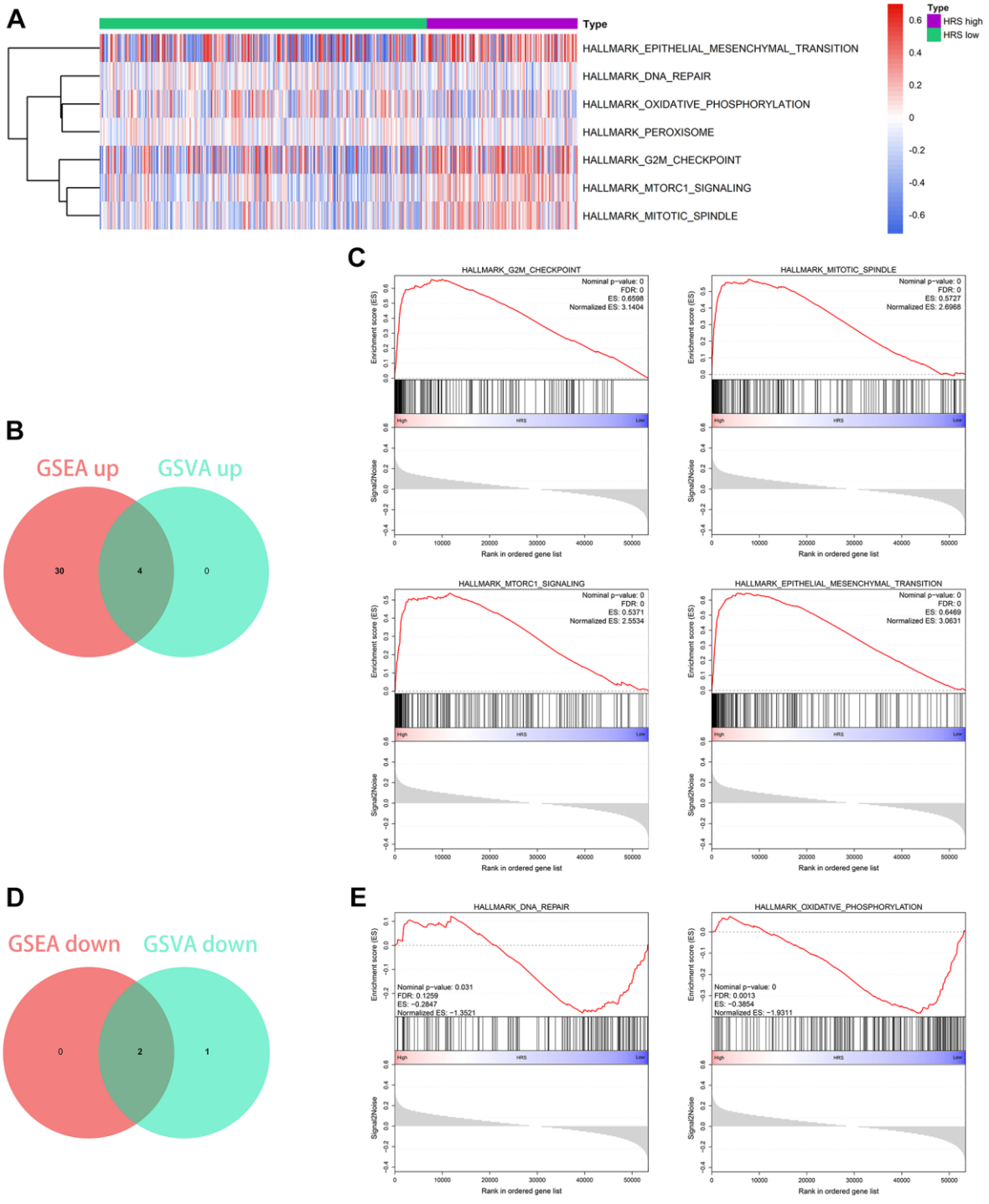


Figure 8. Functional enrichment analysis. (**A**) The heatmap showing the analysis results of GSVA. (**B**, **C**) Four pathways, including G2M checkpoint, mitotic spindle, mTORC1 signaling, and epithelial mesenchymal transition, were positively related to HRS, which was codetermined by GSEA and GSVA. (**D**, **E**) Two pathways, including DNA repair and oxidative phosphorylation, were negatively associated with HRS through the combined analysis by GSEA and GSVA.

would lead to the metastasis depression of melanoma cells by interacting with KAI1, which was a metastasis suppressor [28]. *GPSM3* was found to be up-regulated in normal samples and the cases with favorable prognosis (Supplementary Figures 3–6), indicating *GPSM3* acted as a tumor suppressor for BCa. *GPSM3* was found to be a regulator to immune cells like leukocytes and monocytes [29, 30]. However, the functions of *GPSM3* in malignant tumors remain unclear. Other novel biomarkers, like *RFC4*, *RPA3*, *IFI27L2*, and *APH1A*, were also identified, despite the unclear associated mechanisms. Totally, our risk signature helped identify novel biomarkers and might provide clues for mechanisms of BCa from the prospect of ITH.

The potential compounds targeting HRS were also predicted with CMap. Some of the compounds could interact with malignant tumors from previous research. Lycorine promoted apoptosis of gastric tumor cells by FBXW7-MCL1 axis [31]. Low-dose naltrexone inhibited PI3K/AKT/mTOR pathway and thus suppressed the proliferation of cervical cancer cells [32]. Tanespimycin acted as an antineoplastic drug through targeting HSP90 [33]. Wortmannin was a DNA repair inhibitor and could delay the production of cisplatin resistance [34] in ovarian cancer. Here, we listed some potential compounds targeting ITH, which might help develop some new therapeutic plans.

The shortcomings of the present study should not be neglected. First, although the prognostic value of HRS was validated in 785 BCa cases from three different cohorts, a prospective, large-scale, and multi-center clinical trait would help revalidate the clinical

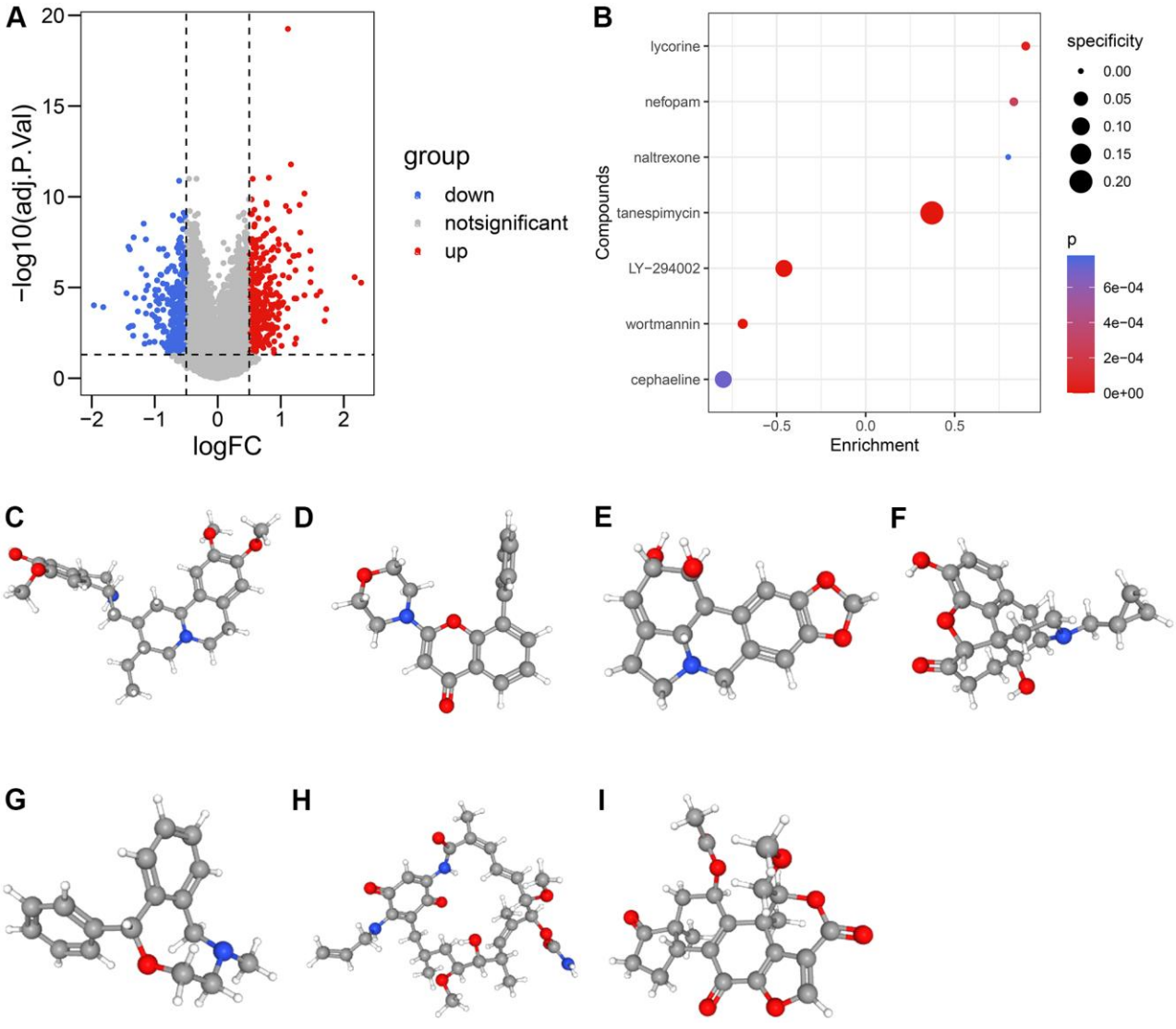


Figure 9. The candidate compounds targeting HRS. (**A**) The volcano plot displayed the DEGs between low-HRS and high-HRS cases. (**B**–**I**) Seven compounds, including cephaeline (**C**), LY-294002 (**D**), lycorine (**E**), naltrexone (**F**), nefopam (**G**), tanespimycin (**H**), and wortmannin (**I**), were identified.

usefulness. Second, several biological pathways and candidate compounds were found to be linked to HRS, which needed to be further experimental validated.

In a word, we developed a novel ITH-related signature to predict the prognosis of BCa based on scRNA-seq, and validated the robustness in large cohorts, which provided a promising tool for clinicians and identified novel biomarkers to disclose the underlying mechanisms.

MATERIALS AND METHODS

Transcriptome data acquisition

The RNA-seq data of 427 samples, containing 19 paracancerous and 408 BCa samples, were downloaded from the TCGA website (https://portal.gdc.cancer.gov/), along with the following-up and clinicopathological information. The external validation datasets, which included GSE13507 and GSE32894, were obtained on the GEO database (https://www.ncbi.nlm.nih.gov/geo/), and their expression matrices were directly downloaded. After excluding the cases with the following-up less than 30 days, a sum of 785 cases, including 396 cases from TCGA, 165 patients from GSE13507 and 224 individuals from GSE32894, were enrolled in this study.

ITH evaluation and TMB calculation

Here, we utilized a reported method, known as the MATH algorithm, to infer the degree of ITH [13, 35]. After obtaining the maf files in the TCGA-BLCA dataset with the TCGAmutations R package, we estimated the MATH value of each individual using the maftools package, which was also implemented to calculate the TMB based on the masked somatic mutation data downloaded from TCGA.

Profiling of scRNA-seq

Compared with the DEGs screened from bulk-seq data, those extracted based on scRNA-seq data might be able to reflect the ITH with higher efficacy. Here, we downloaded the expression matrix of 2075 cells isolated from a patient with primary BCa from GEO database [15]. To make sure the DEGs were all obtained from human, we discarded two other samples in the raw dataset obtained from mice. Seurat package was used to filter the cells with poor quality, where the detected gene, cell, and mitochondria gene counts were also calculated. The cells with less than 200 genes detected and more than 5% of mitochondria gene proportion would be excluded. Based on the Top 1500 variable genes across all cell samples, PCA and t-SNE were then performed to classify the cell samples, and the marker genes were screened with $|logFC| > 0.5$ and adjusted $p < 0.05$ filtering. We annotated the cell categories based on the marker genes with CellMarker [36] (http://biocc.hrbmu.edu.cn/CellMarker/) and CancerSEA [37] (http://biocc.hrbmu.edu.cn/CancerSEA/) databases. Pseudotime analysis was conducted with the monocle package of R.

Survival analysis

Lasso-Cox regression with 10-fold cross-validation and multivariate Cox analysis with stepwise were performed via glmnet and survival packages, respectively. The nomogram and calibration plots were drawn with rms package. Kaplan-Meier survival analysis with log-rank test was carried out with survival R package, where the optimal cut-off value was determined by X-tile [16]. SurvivalROC package helped to conduct the time-dependent ROC analysis.

Gene-pair strategy

To ensure the risk model could suit the BCa samples tested by whatever technical means, a gene-pair strategy was adopted, where there was no need for concrete gene expression values [38]. If the mRNA expression value of gene A is higher than that of gene B, A plus B, or "A|B", is defined as 0; otherwise, it will be regarded as 1. All genes will be cyclically paired, and a 0-or-1 matrix will be successfully constructed. Besides, if 0 or 1 accounts for less than 20% of the gene pairs, this pair would be considered to be meaningless and excluded from the present study.

Evaluation of immune infiltration

To estimate the proportion of immune and stromal components in TME, we implemented the ESTIMATE algorithm [14], which has been widely utilized in many different tumors, including BCa [39]. Besides, the CIBERSORT algorithm was also used for the evaluation of the infiltration level of 22 immune cells [17].

Chemotherapeutic sensitivity estimation

The half inhibitory centration (IC50) of the common chemotherapeutic drugs, like cisplatin, doxorubicin, gemcitabine, methotrexate, and vinblastine, were evaluated with the pRRophetic package on the basis of the transcriptome RNA-seq data.

Functional enrichment analysis

GO and KEGG enrichment was performed through the clusterProfiler package. GSVA was conducted via

GSVA package, and limma package was used for the HRS-related pathways screening under |logFC| > 0.05 and adjusted $p < 0.05$ threshold. GSEA software (version 4.1.0) was downloaded from GSEA's official website (https://www.gsea-msigdb.org/gsea/), and Nom $p < 0.05$ and FDR q < 0.25 were set as the filtering criteria. The hallmark gene signature (version 7.2), downloaded from the Molecular Signatures Database (MSigDB), was chosen as the reference gene set.

Identification of novel candidate compounds

Based on the transcriptome expression data of 396 cases from TCGA, the DEGs extracted between high-HRS and low-HRS groups were screened via limma package. The genes with |logFC| > 0.5 and adjusted $p < 0.05$ were considered to be significant. Subsequently, the gene symbols of the DEGs were all transformed into Affymetrix probe ID, which would be uploaded to the CMap website (https://portals.broadinstitute.org/cmap/). The predicted compounds with $p < 0.001$ would be considered meaningful, and the analysis results were visualized by the ggplot2 package. The 3D structures of the compounds were obtained from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/).

Cell culture and real-time quantitative PCR

T24 and SV-HUC-1 cell lines were purchased from the Chinese Academy of Sciences Shanghai Cell Bank, and cultured in McCoy's 5A modified medium (Gibco, USA) and F12K medium (Gibco, USA), respectively. 1% antibiotic and 10% fetal bovine serum (FBS, Gibco, USA) were added to the medium. The cells were maintained in a humidified atmosphere with 5% CO2 at 37°C.

According to the manufacturer's protocol, the total RNA of these cells was extracted with Trizol (ThermoFisher Scientific, Germany). First-strand cDNA was synthesized with PrimeScript RT Reagent Kit (Takara, China) and amplified by SYBR Premix ExTaq kit (Takara, China). The mRNA expression values were detected via ABI Prism 7000 (Applied Biosystems, USA). GAPDH was chosen as an internal reference, and $2^{-\Delta\Delta C}$ was utilized to calculate the gene expression values. Student's *t* test was used to detect the expression difference. All the primers were obtained from PrimerBank (https://pga.mgh.harvard.edu/primerbank/) and shown in Supplementary Table 12.

AUTHOR CONTRIBUTIONS

CDL designed the whole study and provided financial support. RRZ developed the algorithm and draw the plots. JJL wrote the original draft. QC, HT, and CY did help to editing and reviewing.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This study is supported by National Natural Science Foundation of China (NO. 81772257), Youth Cultivation Program of Southern Medical University (NO. PY2018N076) and Medical Scientific Research Foundation of Guangdong Province (NO. A2019557).

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71:209–49. https://doi.org/10.3322/caac.21660 PMID:33538338

2. Singh DK, Patel VG, Oh WK, Aguirre-Ghiso JA. Prostate Cancer Dormancy and Reactivation in Bone Marrow. J Clin Med. 2021; 10:2648. https://doi.org/10.3390/jcm10122648 PMID:34208521

3. Lenis AT, Lec PM, Chamie K, Mshs MD. Bladder Cancer: A Review. JAMA. 2020; 324:1980–91. https://doi.org/10.1001/jama.2020.17598 PMID:33201207

4. Galsky MD, Stensland KD, Moshier E, Sfakianos JP, McBride RB, Tsao CK, Casey M, Boffetta P, Oh WK, Mazumdar M, Wisnivesky JP. Effectiveness of Adjuvant Chemotherapy for Locally Advanced Bladder Cancer. J Clin Oncol. 2016; 34:825–32. https://doi.org/10.1200/JCO.2015.64.1076 PMID:26786930

5. Powles T. Immunotherapy: The development of immunotherapy in urothelial bladder cancer. Nat Rev Clin Oncol. 2015; 12:193–94. https://doi.org/10.1038/nrclinonc.2015.51 PMID:25781573

6. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. Nat Rev Genet. 2019; 20:404–16. https://doi.org/10.1038/s41576-019-0114-6 PMID:30918367

7. Wolf Y, Bartok O, Patkar S, Eli GB, Cohen S, Litchfield K, Levy R, Jiménez-Sánchez A, Trabish S, Lee JS, Karathia H, Barnea E, Day CP, et al. UVB-Induced Tumor Heterogeneity Diminishes Immune Response in Melanoma. Cell. 2019; 179:219–35.e21.

https://doi.org/10.1016/j.cell.2019.08.032
PMID:31522890

8. Mroz EA, Tward AD, Pickering CR, Myers JN, Ferris RL, Rocco JW. High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. Cancer. 2013; 119:3034–42.
https://doi.org/10.1002/cncr.28150
PMID:23696076

9. Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, Yu K, Rubanova Y, Macintyre G, Demeulemeester J, Vázquez-García I, Kleinheinz K, Livitz DG, et al, and PCAWG Evolution and Heterogeneity Working Group and the PCAWG Consortium. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. Cell. 2021; 184:2239–54.e39.
https://doi.org/10.1016/j.cell.2021.03.009
PMID:33831375

10. Gonzalez-Silva L, Quevedo L, Varela I. Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies. Trends Cancer. 2020; 6:13–19.
https://doi.org/10.1016/j.trecan.2019.11.010
PMID:31952776

11. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, Yu EA, Schenk EL, Tan W, Zee A, Tan M, Gui P, Lea T, et al. Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell. 2020; 182:1232–51.e22.
https://doi.org/10.1016/j.cell.2020.07.017
PMID:32822576

12. Li B, Cui Y, Diehn M, Li R. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. JAMA Oncol. 2017; 3:1529–37.
https://doi.org/10.1001/jamaoncol.2017.1609
PMID:28687838

13. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncol. 2013; 49:211–15.
https://doi.org/10.1016/j.oraloncology.2012.09.007
PMID:23079694

14. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612.
https://doi.org/10.1038/ncomms3612
PMID:24113773

15. Lee HW, Chung W, Lee HO, Jeong DE, Jo A, Lim JE, Hong JH, Nam DH, Jeong BC, Park SH, Joo KM, Park WY. Single-cell RNA sequencing reveals the tumor microenvironment and facilitates strategic choices to circumvent treatment failure in a chemorefractory bladder cancer patient. Genome Med. 2020; 12:47.
https://doi.org/10.1186/s13073-020-00741-6
PMID:32460812

16. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2004; 10:7252–59.
https://doi.org/10.1158/1078-0432.CCR-04-0713
PMID:15534099

17. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015; 12:453–57.
https://doi.org/10.1038/nmeth.3337
PMID:25822800

18. Maj T, Wang W, Crespo J, Zhang H, Wang W, Wei S, Zhao L, Vatan L, Shao I, Szeliga W, Lyssiotis C, Liu JR, Kryczek I, Zou W. Oxidative stress controls regulatory T cell apoptosis and suppressor activity and PD-L1-blockade resistance in tumor. Nat Immunol. 2017; 18:1332–41.
https://doi.org/10.1038/ni.3868
PMID:29083399

19. Yang G, Ni JS, Li Y, Zha M, Tu Y, Li K. Acceptor Engineering for Optimized ROS Generation Facilitates Reprogramming Macrophages to M1 Phenotype in Photodynamic Immunotherapy. Angew Chem Int Ed Engl. 2021; 60:5386–93.
https://doi.org/10.1002/anie.202013228
PMID:33236483

20. Somasundaram R, Connelly T, Choi R, Choi H, Samarkina A, Li L, Gregorio E, Chen Y, Thakur R, Abdel-Mohsen M, Beqiri M, Kiernan M, Perego M, et al. Tumor-infiltrating mast cells are associated with resistance to anti-PD-1 therapy. Nat Commun. 2021; 12:346.
https://doi.org/10.1038/s41467-020-20600-7
PMID:33436641

21. Kulkarni GS, Black PC, Sridhar SS, Kapoor A, Zlotta AR, Shayegan B, Rendon RA, Chung P, van der Kwast T, Alimohamed N, Fradet Y, Kassouf W. Canadian Urological Association guideline: Muscle-invasive bladder cancer. Can Urol Assoc J. 2019; 13:230–38.
https://doi.org/10.5489/cuaj.5902
PMID:30763236

22. Sun XX, Yu Q. Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. Acta Pharmacol Sin. 2015; 36:1219–27.
https://doi.org/10.1038/aps.2015.92
PMID:26388155

23. Kondo H, Ratcliffe CDH, Hooper S, Ellis J, MacRae JI, Hennequart M, Dunsby CW, Anderson KI, Sahai E. Single-cell resolved imaging reveals intra-tumor heterogeneity in glycolysis, transitions between metabolic states, and their regulatory mechanisms. Cell Rep. 2021; 34:108750. https://doi.org/10.1016/j.celrep.2021.108750 PMID:33596424

24. Wu X, Lv D, Cai C, Zhao Z, Wang M, Chen W, Liu Y. A TP53-Associated Immune Prognostic Signature for the Prediction of Overall Survival and Therapeutic Responses in Muscle-Invasive Bladder Cancer. Front Immunol. 2020; 11:590618. https://doi.org/10.3389/fimmu.2020.590618 PMID:33391264

25. Liu Z, Sun T, Zhang Z, Bi J, Kong C. An 18-gene signature based on glucose metabolism and DNA methylation improves prognostic prediction for urinary bladder cancer. Genomics. 2021; 113:896–907. https://doi.org/10.1016/j.ygeno.2020.10.022 PMID:33096258

26. Su M, Qiao KY, Xie XL, Zhu XY, Gao FL, Li CJ, Zhao DQ. Development of a Prognostic Signature Based on Single-Cell RNA Sequencing Data of Immune Cells in Intrahepatic Cholangiocarcinoma. Front Genet. 2021; 11:615680. https://doi.org/10.3389/fgene.2020.615680 PMID:33613623

27. Zhang C, He H, Hu X, Liu A, Huang D, Xu Y, Chen L, Xu D. Development and validation of a metastasis-associated prognostic signature based on single-cell RNA-seq in clear cell renal cell carcinoma. Aging (Albany NY). 2019; 11:10183–202. https://doi.org/10.18632/aging.102434 PMID:31747386

28. Marie KL, Sassano A, Yang HH, Michalowski AM, Michael HT, Guo T, Tsai YC, Weissman AM, Lee MP, Jenkins LM, Zaidi MR, Pérez-Guijarro E, Day CP, et al. Melanoblast transcriptome analysis reveals pathways promoting melanoma metastasis. Nat Commun. 2020; 11:333. https://doi.org/10.1038/s41467-019-14085-2 PMID:31949145

29. Billard MJ, Gall BJ, Richards KL, Siderovski DP, Tarrant TK. G protein signaling modulator-3: a leukocyte regulator of inflammation in health and disease. Am J Clin Exp Immunol. 2014; 3:97–106. PMID:25143870

30. Giguere PM, Billard MJ, Laroche G, Buckley BK, Timoshchenko RG, McGinnis MW, Esserman D, Foreman O, Liu P, Siderovski DP, Tarrant TK. G-protein signaling modulator-3, a gene linked to autoimmune diseases, regulates monocyte function and its deficiency protects from inflammatory arthritis. Mol Immunol. 2013; 54:193–98. https://doi.org/10.1016/j.molimm.2012.12.001 PMID:23280397

31. Li C, Deng C, Pan G, Wang X, Zhang K, Dong Z, Zhao G, Tan M, Hu X, Shi S, Du J, Ji H, Wang X, et al. Lycorine hydrochloride inhibits cell proliferation and induces apoptosis through promoting FBXW7-MCL1 axis in gastric cancer. J Exp Clin Cancer Res. 2020; 39:230. https://doi.org/10.1186/s13046-020-01743-3 PMID:33126914

32. Liu N, Yan L, Shan F, Wang X, Qu N, Handley MK, Ma M. Low-dose naltrexone plays antineoplastic role in cervical cancer progression through suppressing PI3K/AKT/mTOR pathway. Transl Oncol. 2021; 14:101028. https://doi.org/10.1016/j.tranon.2021.101028 PMID:33540155

33. Liu J, Meng H, Li S, Shen Y, Wang H, Shan W, Qiu J, Zhang J, Cheng W. Identification of Potential Biomarkers in Association With Progression and Prognosis in Epithelial Ovarian Cancer by Integrated Bioinformatics Analysis. Front Genet. 2019; 10:1031. https://doi.org/10.3389/fgene.2019.01031 PMID:31708970

34. Zhang M, Hagan CT 4th, Min Y, Foley H, Tian X, Yang F, Mi Y, Au KM, Medik Y, Roche K, Wagner K, Rodgers Z, Wang AZ. Nanoparticle co-delivery of wortmannin and cisplatin synergistically enhances chemoradiotherapy and reverses platinum resistance in ovarian cancer models. Biomaterials. 2018; 169:1–10. https://doi.org/10.1016/j.biomaterials.2018.03.055 PMID:29631163

35. Ma D, Jiang YZ, Liu XY, Liu YR, Shao ZM. Clinical and molecular relevance of mutant-allele tumor heterogeneity in breast cancer. Breast Cancer Res Treat. 2017; 162:39–48. https://doi.org/10.1007/s10549-017-4113-z PMID:28093659

36. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. 2019; 47:D721–28. https://doi.org/10.1093/nar/gky900 PMID:30289549

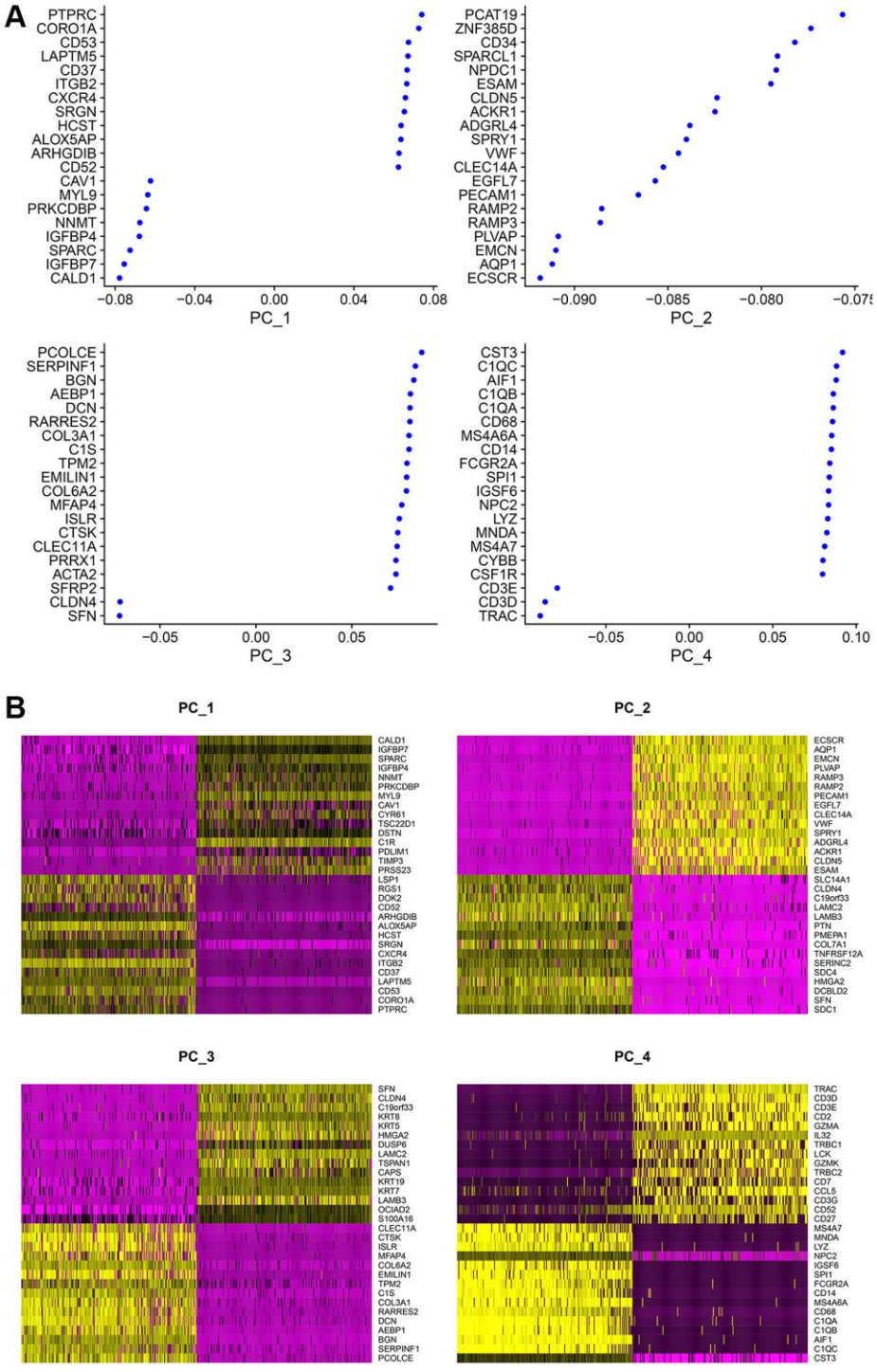
37. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, Xu L, Luo T, Yan H, Long Z, Shi A, Zhao T, Xiao Y, Li X. CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res. 2019; 47:D900–08. https://doi.org/10.1093/nar/gky939 PMID:30329142

38. Hong W, Liang L, Gu Y, Qi Z, Qiu H, Yang X, Zeng W, Ma L, Xie J. Immune-Related lncRNA to Construct Novel Signature and Predict the Immune Landscape of Human Hepatocellular Carcinoma. Mol Ther Nucleic Acids. 2020; 22:937–47. https://doi.org/10.1016/j.omtn.2020.10.002 PMID:33251044

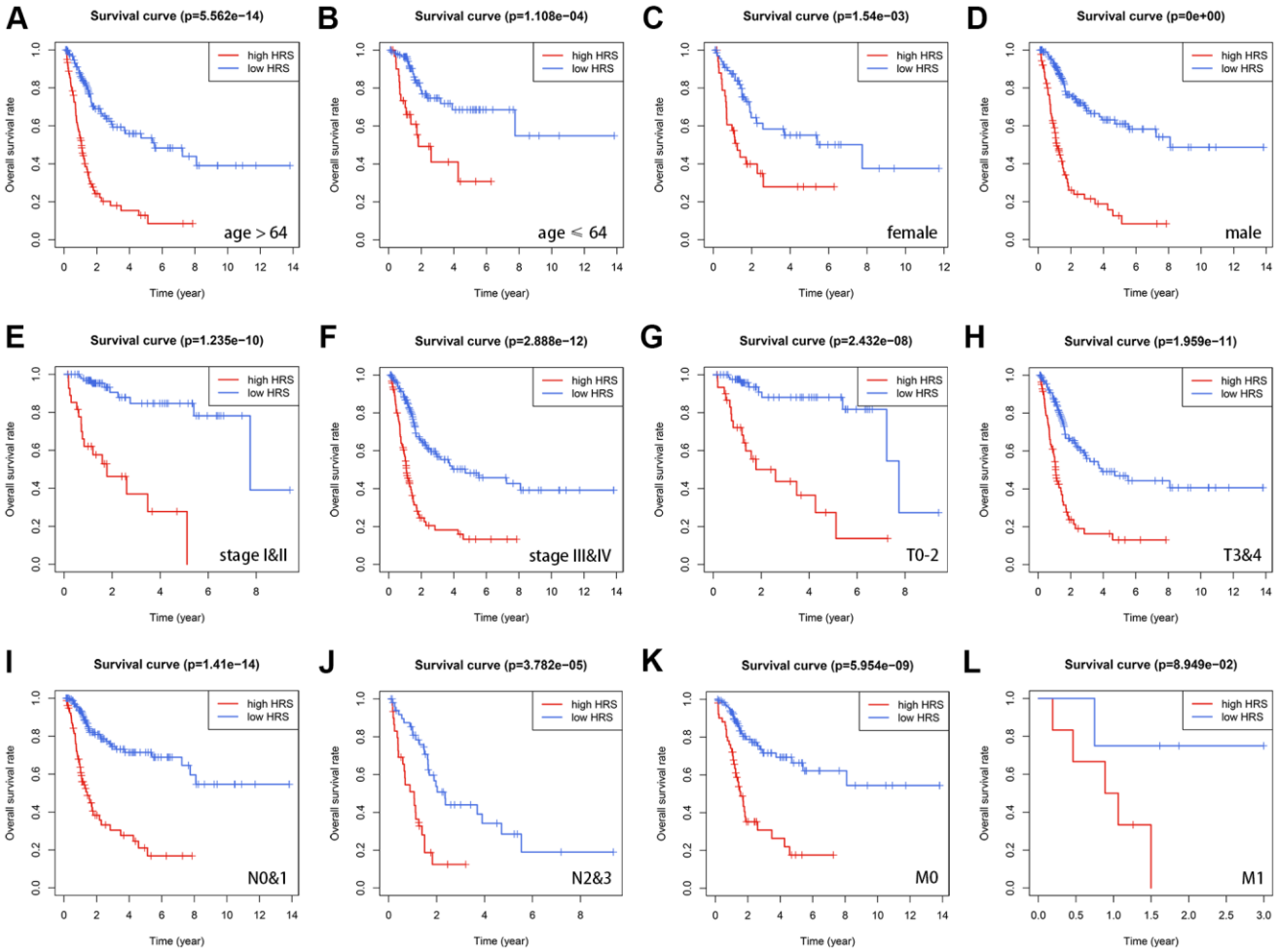
39. Cao R, Yuan L, Ma B, Wang G, Tian Y. Immune-related long non-coding RNA signature identified prognosis and immunotherapeutic efficiency in bladder cancer (BLCA). Cancer Cell Int. 2020; 20:276. https://doi.org/10.1186/s12935-020-01362-0 PMID:32607061

SUPPLEMENTARY MATERIALS

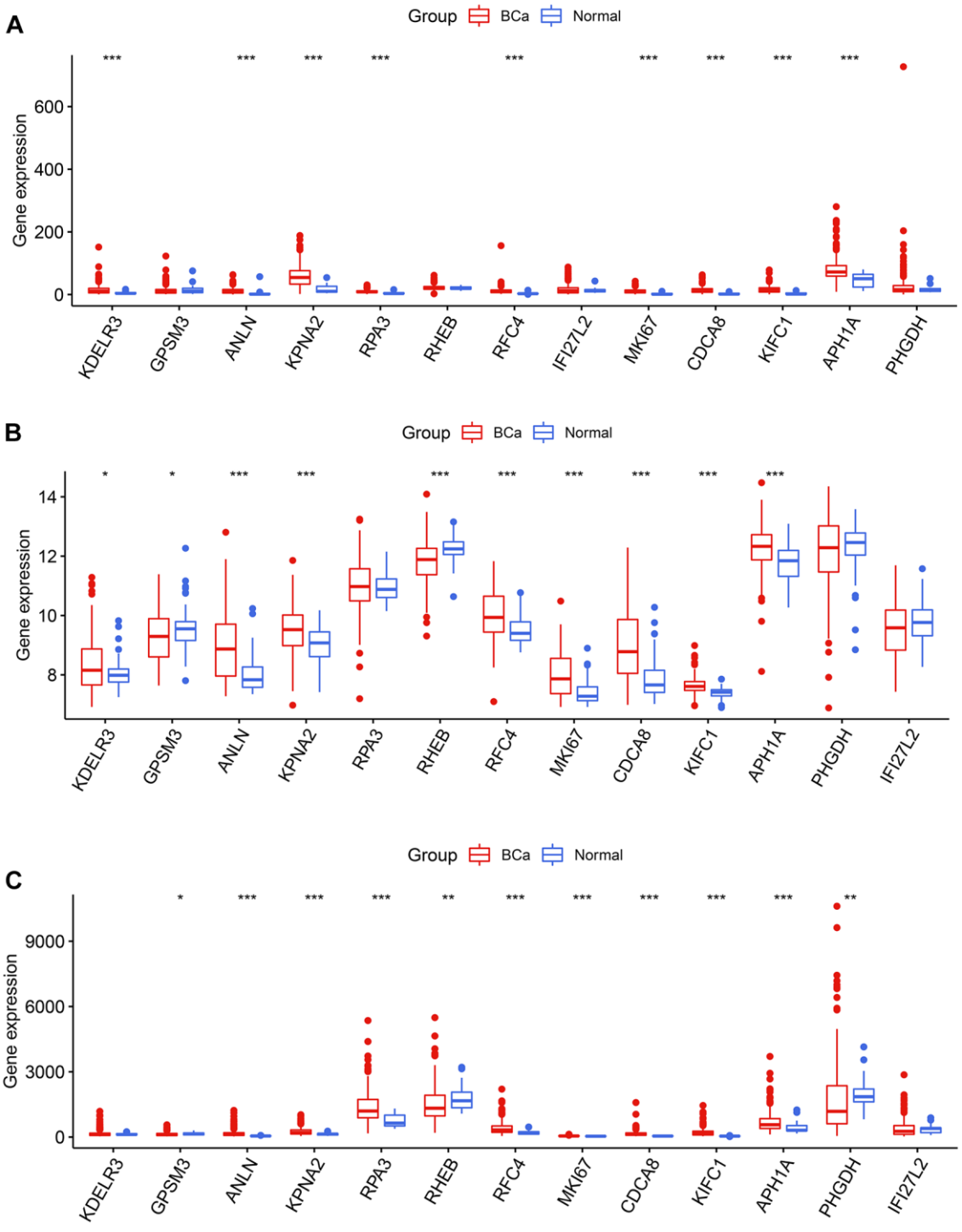
Supplementary Figures



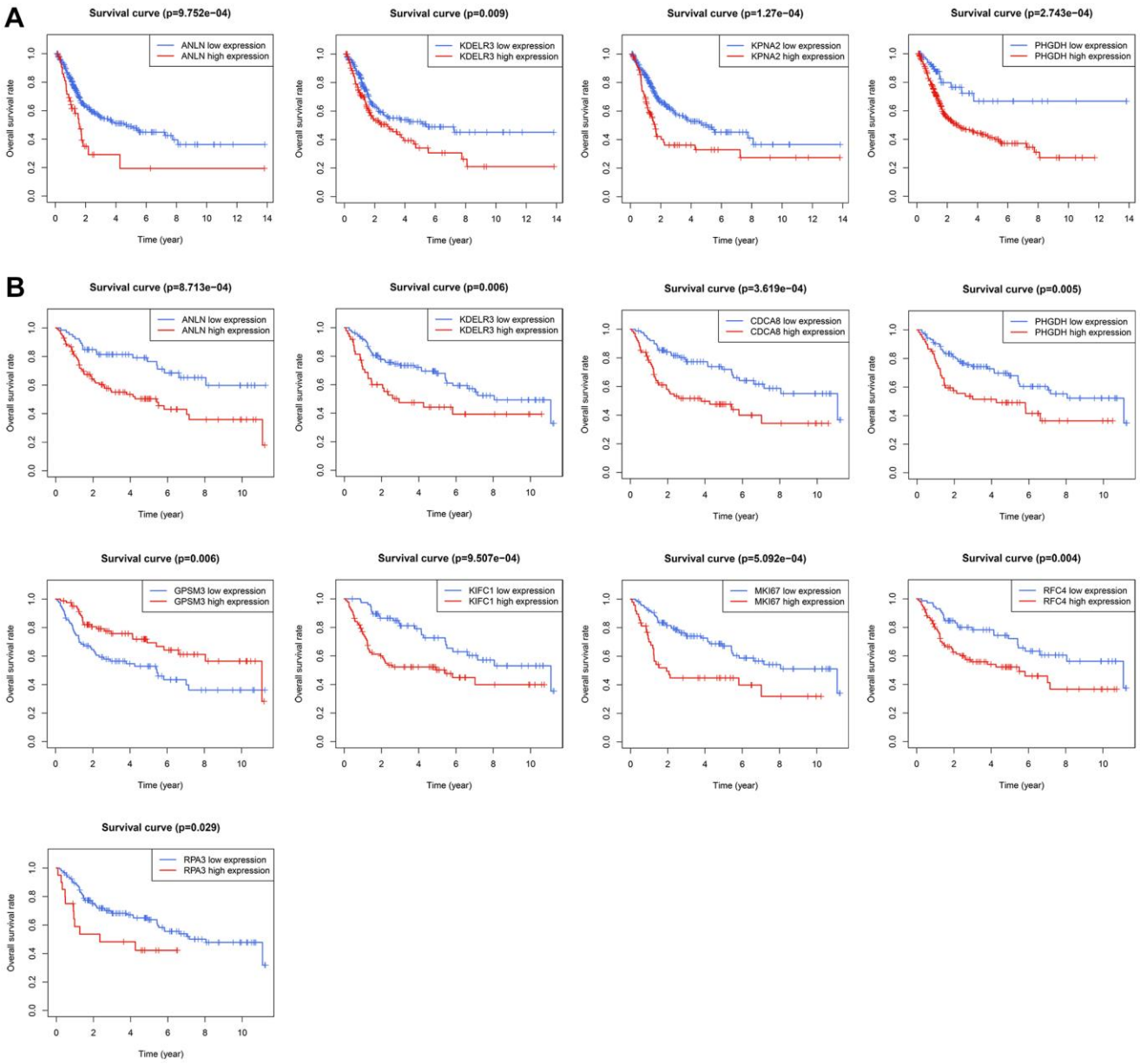
Supplementary Figure 1. The Top 4 components and the correlated genes in PCA analysis. (**A**) The Top related genes to each principal component. (**B**) The heatmap indicating the expression level of the Top related genes. The colors ranging from purple to yellow represented the expression values from low to high.



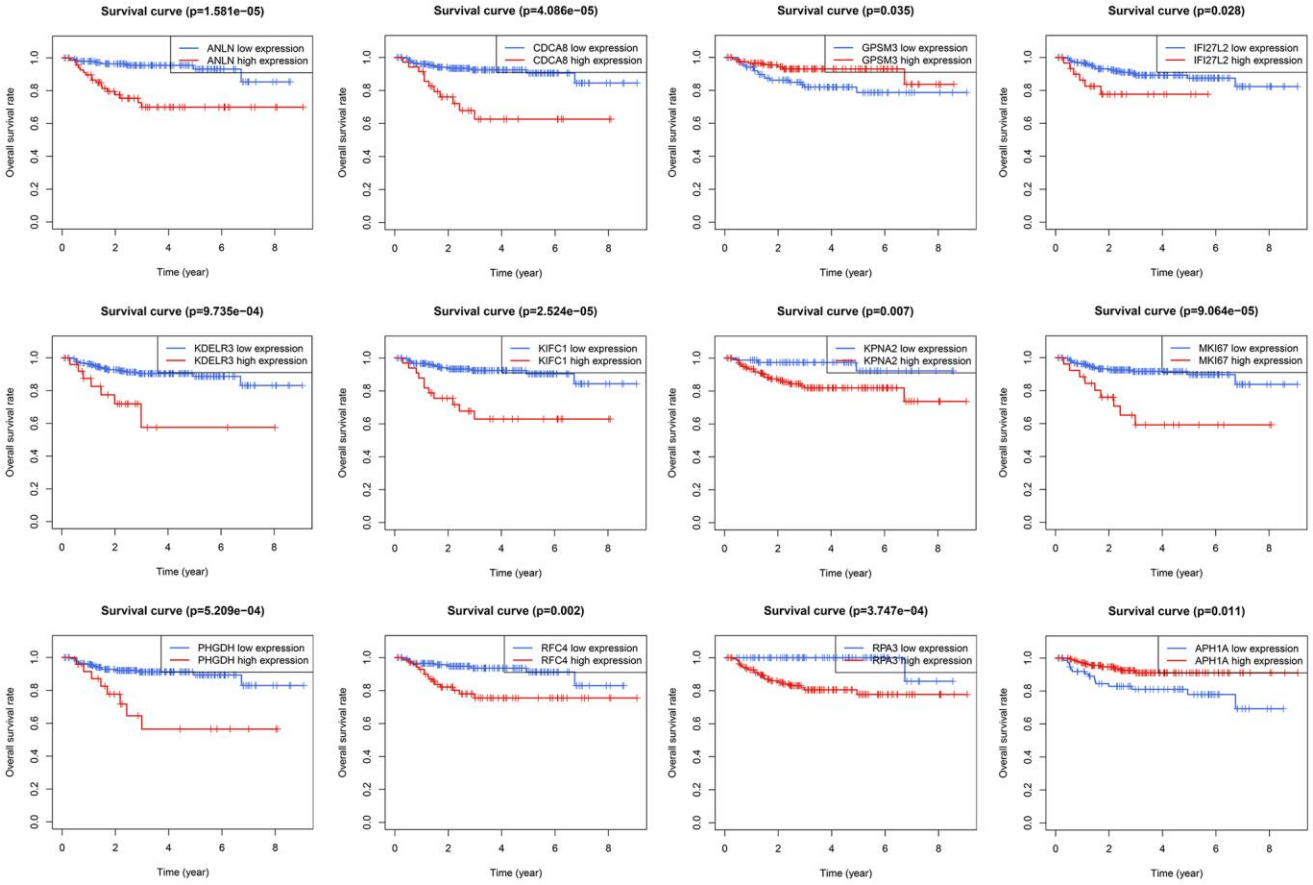
Supplementary Figure 2. The Kaplan-Meier survival stratification analysis in TCGA-BLCA cohorts. (**A**) Age > 64 years. (**B**) Age ≤ 64 years. (**C**) Female. (**D**) Male. (**E**) Pathological stage I and II. (**F**) Pathological stage III and IV. (**G**) Pathological T0-2. (**H**) Pathological T3 and 4. (**I**) Pathological N0 and 1. (**J**) Pathological N2 and 3. (**K**) M0. (**L**) M1.



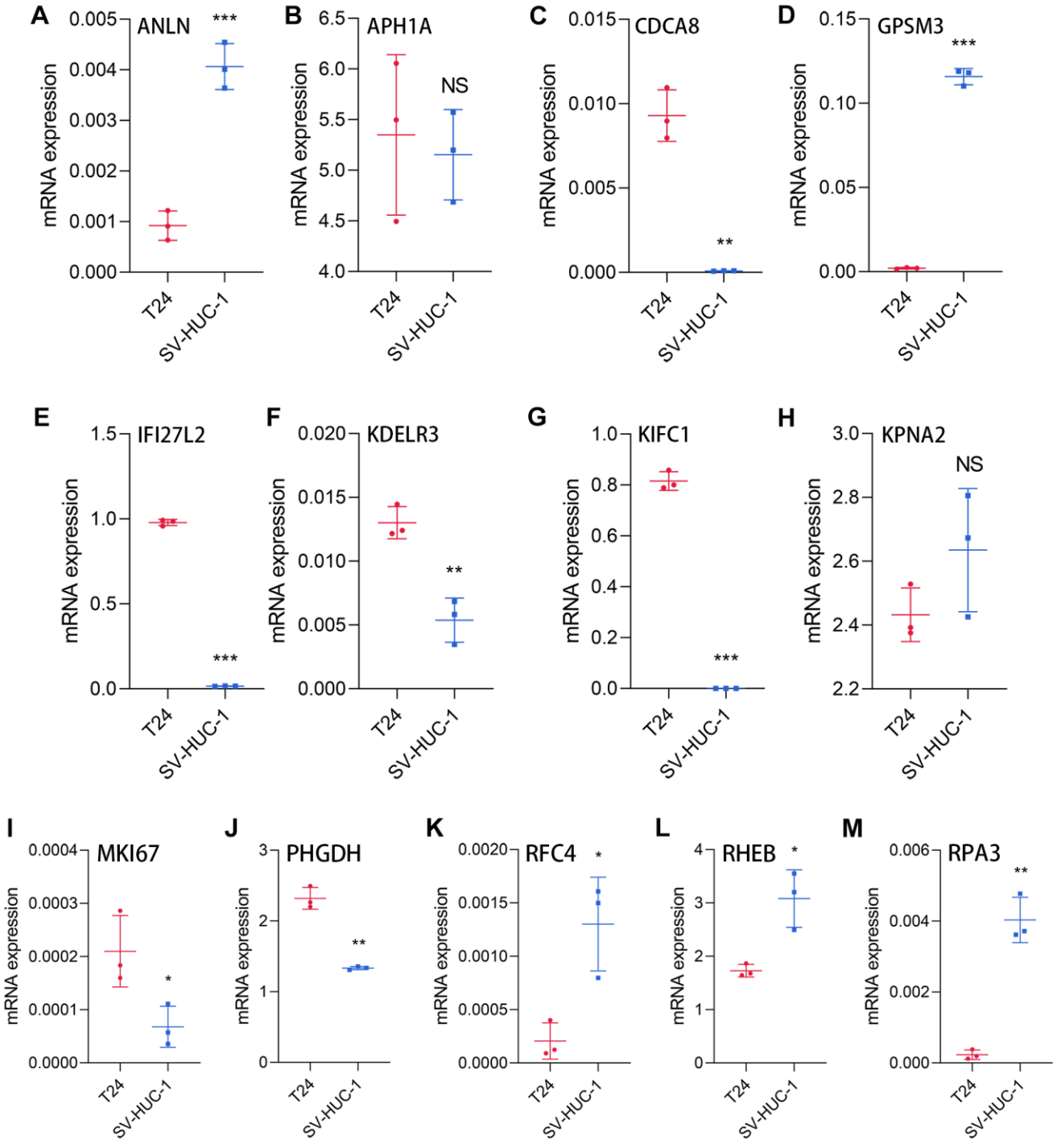
Supplementary Figure 3. The expression level of 13 genes comprising HRS in paracancerous and BCa tissues in TCGA (**A**), GSE13507 (**B**), and GSE32894 (**C**) cohorts. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.



Supplementary Figure 4. Kaplan-Meier survival analysis of the 13 genes in TCGA (**A**) and GSE13507 (**B**) datasets. The optimal cut-off was determined through X-tile.



Supplementary Figure 5. Kaplan-Meier survival analysis of the 13 genes in GSE32894 cohort. X-tile was used for detecting the optimal cut-off value.



Supplementary Figure 6. The expression values of the 13 HRS genes, including ANLN (**A**), APH1A (**B**), CDCA8 (**C**), GPSM3 (**D**), IFI27L2 (**E**), KDELR3 (**F**), KIFC1 (**G**), KPNA2 (**H**), MKI67 (**I**), PHGDH (**J**), RFC4 (**K**), RHEB (**L**), and RPA3 (**M**), in T24 cells and SV-HUC-1 cells via RT-qPCR detection. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–4, 6–8, and 14–17.

Supplementary Table 1. The evaluated MATH value of each individual from TCGA-BLCA cohort.

Supplementary Table 2. 2075 cell samples were classified into 14 clusters.

Supplementary Table 3. The screened cell markers of 14 cell clusters.

Supplementary Table 4. The spearman correlation coefficients of 2940 cell markers to MATH values.

Supplementary Table 5. The details of the established model constructed via multivariate Cox regression.

Id	coef	HR	HR.95L	HR.95H	*p* value
KDELR3_RFC4	0.59906044	1.820407615	1.284993084	2.578911845	0.000749008
GPSM3_IFI27L2	–0.437820457	0.645441657	0.452983804	0.919668494	0.015371371
ANLN_MKI67	0.374716571	1.454579086	1.0085688	2.097824479	0.0448955
ANLN_CDCA8	0.589653534	1.803363503	1.199952669	2.710206833	0.004554417
ANLN_KIFC1	0.343999231	1.410577551	0.946384138	2.102453906	0.091154079
KPNA2_APH1A	0.501950219	1.651939776	1.164418355	2.343577815	0.004907379
RPA3_PHGDH	–0.623979584	0.535807896	0.324557473	0.884558593	0.01470563
RHEB_PHGDH	–0.41459758	0.660606069	0.464862245	0.938773546	0.020758688

Supplementary Table 6. The HRS and risk grouping of the cases from TCGA-BLCA cohort.

Supplementary Table 7. The HRS and risk grouping of the patients from GSE13507 cohort.

Supplementary Table 8. The HRS and risk grouping of the patients from GSE32894 cohort.

Supplementary Table 9. The *p* values of the 13 genes to OS via Kaplan-Meier survival analysis with log-rank test in TCGA cohort.

Gene	*p* value
KDELR3	0.0089878
GPSM3	0.0552757
ANLN	0.0009752
KPNA2	0.000127
RPA3	0.1100399
RHEB	0.3161838
RFC4	0.3074479
IFI27L2	0.1198286
MKI67	0.1708345
CDCA8	0.2304972
KIFC1	0.5828702
APH1A	0.0906864
PHGDH	0.0002743

Supplementary Table 10. The *p* values of the 13 genes to OS via Kaplan-Meier survival analysis with log-rank test in GSE13507 cohort.

Gene	*p* value
KDELR3	0.0058833
GPSM3	0.0062059
ANLN	0.0008713
KPNA2	0.1358367
RPA3	0.0288977
RHEB	0.1812605
RFC4	0.0043104
MKI67	0.0005092
CDCA8	0.0003619
KIFC1	0.0009507
APH1A	0.1464289
PHGDH	0.0050507
IFI27L2	0.0824967

Supplementary Table 11. The *p* values of the 13 genes to OS via Kaplan-Meier survival analysis with log-rank test in GSE32894 cohort.

Gene	*p* value
KDELR3	0.000973
GPSM3	0.035384
ANLN	1.58E-05
KPNA2	0.00709
RPA3	0.000375
RHEB	0.331377
RFC4	0.002442
MKI67	9.06E-05
CDCA8	4.09E-05
KIFC1	2.52E-05
APH1A	0.011216
PHGDH	0.000521
IFI27L2	0.028095

Supplementary Table 12. The primers utilized in present study.

Id	Forward primer (5′ –> 3′)	Reverse primer (5′ –> 3′)
KDELR3	TCCCAGTCATTGGCCTTTCC	CCAGTTAGCCAGGTAGAGTGC
GPSM3	AGGAGTTTTTCCCAGTCTCAGT	TTCTCTTCCCACCCAAACAGC
ANLN	TGCCAGGCGAGAGAATCTTC	CGCTTAGCATGAGTCATAGACCT
KPNA2	CTGCCCGTCTTCACAGATTCA	GCGGAGAAGTAGCATCATCAGG
RPA3	AGCTCAATTCATCGACAAGCC	TCTTCATCAAGGGGTTCCATCA
RHEB	TTGTGGACTCCTACGATCCAA	GGCTGTGTCTACAAGTTGAAGAT
RFC4	CCGCTGACCAAGGATCGAG	AGGGAACGGGTTTGGCTTTC
IFI27L2	CCACATCATCCAACATCCTCC	TCATCTTCTTTAGCCTCGGGTT

MKI67	ACGCCTGGTTACTATCAAAAGG	CAGACCCATTTACTTGTGTTGGA
CDCA8	GAAGGGCAGTAGTCGGGTG	TCACGGTCGAAGTCTTTCAGA
KIFC1	GGTGCAACGACCAAAATTACC	GGGTCCTGTCTTCTTGGAAAC
APH1A	TTTTTCGGCTGCACTTTCGTC	TGCGACCAGGATGATAACGC
PHGDH	CTGCGGAAAGTGCTCATCAGT	TGGCAGAGCGAACAATAAGGC

Supplementary Table 13. The results of GSVA analysis.

Id	logFC	AveExpr	*t*	*p* value	adj. *P*. Val	B
HALLMARK_MTORC1_SIGNALING	0.091526202	−0.04214319	4.418777	1.28E-05	0.0006389	2.8557429
HALLMARK_OXIDATIVE_PHOSPHORYLATION	−0.102512372	−0.04601685	–4.206509	3.20E-05	0.0015677	2.0000334
HALLMARK_MITOTIC_SPINDLE	0.090202864	−0.04325266	4.1063592	4.87E-05	0.0023374	1.6096638
HALLMARK_DNA_REPAIR	−0.0680874	−0.04970153	–3.971025	8.48E-05	0.0039839	1.0958871
HALLMARK_PEROXISOME	−0.059440425	−0.03764865	–3.824692	0.0001517	0.006978	0.5582572
HALLMARK_G2M_CHECKPOINT	0.12383884	−0.02058956	3.6655883	0.0002798	0.012592	−0.005001
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	0.136256246	−0.01514403	3.5184996	0.0004835	0.0212741	−0.50582
HALLMARK_E2F_TARGETS	0.11558866	−0.01157635	3.0847107	0.002178	0.0936557	−1.869989
HALLMARK_FATTY_ACID_METABOLISM	−0.045102033	−0.04568334	–2.986091	0.0029983	0.1259277	−2.156347
HALLMARK_PROTEIN_SECRETION	0.068457724	−0.03805288	2.9706073	0.0031502	0.129157	−2.200499
HALLMARK_ANGIOGENESIS	0.092499901	−0.025644	2.918166	0.0037184	0.1487362	−2.348405
HALLMARK_ANDROGEN_RESPONSE	0.054163091	−0.04380175	2.8860864	0.0041106	0.1603138	−2.437639
HALLMARK_HYPOXIA	0.057318158	−0.0384402	2.8665723	0.0043673	0.165956	−2.491458
HALLMARK_BILE_ACID_METABOLISM	−0.043584756	−0.04440084	–2.837239	0.0047806	0.1768806	−2.571699
HALLMARK_ADIPOGENESIS	−0.038230578	−0.04633641	–2.596801	0.0097542	0.3511514	−3.199459
HALLMARK_UV_RESPONSE_DN	0.057263757	−0.04566087	2.5444044	0.011319	0.3961639	−3.329144
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	0.043122271	−0.04807679	2.476172	0.0136901	0.465464	−3.494188
HALLMARK_SPERMATOGENESIS	0.035904577	−0.02105281	2.4301759	0.0155276	0.5124106	−3.602991
HALLMARK_TGF_BETA_SIGNALING	0.055378448	−0.04224624	2.3295063	0.020326	0.6504325	−3.834211
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	−0.04290982	−0.05314658	–2.27996	0.0231323	0.7170998	−3.944516
HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.070965528	−0.03196946	2.2615103	0.0242604	0.7278121	−3.985
HALLMARK_COAGULATION	0.056094838	−0.01821846	2.2064716	0.0279148	0.8095302	−4.103868
HALLMARK_HEDGEHOG_SIGNALING	0.052376405	−0.04787308	2.205812	0.0279614	0.8095302	−4.105275
HALLMARK_GLYCOLYSIS	0.0313427	−0.04681433	2.161573	0.0312398	0.8434745	−4.198719
HALLMARK_INFLAMMATORY_RESPONSE	0.069240648	−0.02762496	2.0932041	0.0369561	0.9608593	−4.339494
HALLMARK_MYC_TARGETS_V1	0.05896169	−0.03787073	2.0074322	0.0453714	1	−4.509842
HALLMARK_APICAL_JUNCTION	0.043373326	−0.04002215	2.006346	0.0454876	1	−4.511954
HALLMARK_XENOBIOTIC_METABOLISM	−0.027357847	−0.03182934	–1.875082	0.061505	1	−4.758987
HALLMARK_COMPLEMENT	0.043843668	−0.03093177	1.6110832	0.1079452	1	−5.206029
HALLMARK_KRAS_SIGNALING_UP	0.039222297	−0.03820804	1.5973272	0.1109772	1	−5.227493
HALLMARK_MYC_TARGETS_V2	0.047231895	−0.03622899	1.4396179	0.1507525	1	−5.460575
HALLMARK_IL6_JAK_STAT3_SIGNALING	0.044389386	−0.02740203	1.35617	0.1758051	1	−5.574214
HALLMARK_APOPTOSIS	0.022787978	−0.03955432	1.1226794	0.2622427	1	−5.856423
HALLMARK_NOTCH_SIGNALING	−0.020422791	−0.04917705	–1.120955	0.2629753	1	−5.858312
HALLMARK_IL2_STAT5_SIGNALING	0.025344043	−0.04030152	1.0703473	0.2851042	1	−5.912431
HALLMARK_UV_RESPONSE_UP	−0.012238878	−0.04667981	–0.906738	0.3650877	1	−6.070377
HALLMARK_P53_PATHWAY	−0.012958715	−0.04582705	–0.739255	0.460183	1	−6.205088

HALLMARK_PANCREAS_BETA_CELLS	0.011969347	0.002796107	0.607188	0.5440684	1	−6.292031
HALLMARK_HEME_METABOLISM	0.0066419	−0.04233163	0.5188347	0.604161	1	−6.340692
HALLMARK_ALLOGRAFT_REJECTION	0.013818118	−0.02391175	0.3724455	0.7097571	1	−6.404527
HALLMARK_KRAS_SIGNALING_DN	−0.005060721	−0.0324103	–0.366763	0.7139888	1	−6.406582
HALLMARK_INTERFERON_GAMMA_RESPONSE	0.01207391	−0.02766363	0.3239049	0.7461782	1	−6.421068
HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.005372206	−0.04847371	0.3000273	0.7643112	1	−6.428358
HALLMARK_ESTROGEN_RESPONSE_EARLY	0.003545514	−0.04628076	0.2088272	0.8346888	1	−6.451071
HALLMARK_INTERFERON_ALPHA_RESPONSE	−0.007399147	−0.02553256	–0.18075	0.8566546	1	−6.456425
HALLMARK_APICAL_SURFACE	0.003052593	−0.04552941	0.161198	0.8720183	1	−6.459698
HALLMARK_ESTROGEN_RESPONSE_LATE	−0.002251102	−0.04602582	–0.1469	0.8832846	1	−6.461854
HALLMARK_MYOGENESIS	0.002295383	−0.03740793	0.0912882	0.927309	1	−6.468338
HALLMARK_WNT_BETA_CATENIN_SIGNALING	0.001386947	−0.05009891	0.0732182	0.9416688	1	−6.469793
HALLMARK_PI3K_AKT_MTOR_SIGNALING	0.000522011	−0.04980514	0.0321418	0.9743749	1	−6.471911

Supplementary Table 14. The results of GSEA analysis for the cases with high-HRS in TCGA-BLCA dataset.

Supplementary Table 15. The results of GSEA analysis for the cases with low-HRS in TCGA-BLCA dataset.

Supplementary Table 16. The differentially-expressed genes between high-HRS and low-HRS groups through limma.

Supplementary Table 17. The potential compounds targeting HRS via Connectivity Map analysis.