

Long non-coding RNAs to predict postoperative recurrence in muscle-invasive bladder cancer and to develop a new molecular classification system

Zhiyong Li (✉ lizhiy@sysucc.org.cn)

Sun Yat-sen University Cancer Center <https://orcid.org/0000-0001-5225-9012>

LIJUAN JIANG

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

ZHILING ZHANG

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

MINHUA DENG

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

WENSU WEI

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

HUANCHENG TANG

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

SHENGJIE GUO

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

YUNLIN YE

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

KAI YAO

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

ZHUOWEI LIU

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

FANGJIAN ZHOU

Sun Yat-sen University Affiliated Tumor Hospital: Sun Yat-sen University Cancer Center

Primary research

Keywords: biomarker, lncRNA, bladder cancer, recurrence, molecular subtypes

Posted Date: January 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-153937/v1>

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

[Read Full License](#)

Abstract

Background: Reliable molecular markers are much needed for early prediction of recurrence in muscle-invasive bladder cancer (MIBC) patients. We aimed to build an lncRNA signature to improve recurrence prediction and lncRNA-based molecular classification of MIBC.

Methods: lncRNAs of 320 MIBC patients from TCGA database were analyzed, and a nomogram was established. A molecular classification system was created, and immunotherapy, chemotherapy response prediction, immune score analysis, immunoinfiltration analysis, and mutational data analysis were conducted.

Results: An eight-lncRNA signature could classify the patients into high- and low-risk subgroups having significantly different DFS. Our samples validated that the 8 lncRNAs were related with DFS. The nomogram achieved a C-index of 0.719 (95% CI, 0.674–0.764). Time-dependent ROC analyses indicated the superior prognostic accuracy of nomograms for DFS prediction (0.76, 95% CI, 0.697–0.807). Further, four clusters with different DFS and molecular features were identified. The four clusters (median DFS = 11.8, 15.3, 17.9, and 18.9 months, respectively) showed a high frequency of TTN, fibroblast growth factor receptor-3, TP53, and TP53 mutations, respectively. They were enriched with macrophages M2, T cells CD8, macrophages M0, and macrophages M0, respectively. Clusters 2 and 3 demonstrated potential sensitivity to immunotherapy and insensitivity to chemotherapy, whereas cluster 4 showed potential insensitivity to immunotherapy and sensitivity to chemotherapy.

Conclusion: Our lncRNA-based signature proved to be a reliable prognostic tool to determine postoperative recurrence in MIBC patients. We believe that the four molecular subtypes possess the potential to guide treatment.

Background

Bladder cancer is one of the most frequently diagnosed cancers globally and the second most common cancer of the genitourinary tract [1, 2]. Muscle-invasive bladder cancer (MIBC) is associated with a 5-year survival rate of 60% in case of patients with localized disease and < 10% in case of those with distant metastases. Radical cystectomy (RC) is the standard method of treatment. However, a “surgery only” study reported that the 5-year recurrence-free survival was 74%, 52%, and 36% in case of patients with pT2, pT3, and pT4, respectively [3]. More than 20% patients with bladder cancer experience postoperative recurrence, leading to poor survival. Although the use of adjuvant chemotherapy after RC for patients with pT3/4 and/or LN-positive disease without clinically detectable metastases is still under debate, it is effective for treating patients with MIBC with a high risk of recurrence. However, clinicopathological risk factors and TNM stage cannot adequately classify patients into a high or low risk of disease recurrence, and thus, they cannot indicate which patients are likely to benefit from postoperative chemotherapy. Considering this issue, there is an urgent need to identify novel, reliable recurrence-associated molecular

biomarkers that allow for better prognostic stratification and more appropriate therapies for patients with bladder cancer.

MIBC is a heterogeneous condition that is characterized by genomic instability and a high mutation rate. Bladder cancer can be stratified into different molecular subtypes for predicting outcome and treatment response; however, most previous studies have focused on the expression levels of mRNA to differentiate among molecular subtypes.

Long non-coding RNAs (lncRNAs) lack the protein-coding function and are usually > 200 bp in length [3, 4]. Their aberrant expression has been closely related with various types of cancers; moreover, they have been experimentally validated to be involved in the etiology, pathogenesis, and progression of cancers [5]. Furthermore, they evidently are a novel biomarker and therapeutic target for bladder cancer [6]. However, the potential of lncRNA-based signatures and using lncRNAs for the molecular classification of bladder cancer has not been sufficiently explored as yet.

In this study, we aimed to assess the ability of lncRNA expression profiles to predict disease recurrence; further, we attempted to use lncRNAs for the molecular classification of MIBC so as to guide clinical treatment in patients with MIBC.

Materials And Methods

Patients with MIBC

We downloaded TCGA–BLCA RNA sequencing dataset and corresponding clinical characteristics of patients from TCGA website (<https://cancer.genome.nih.gov/>). In total, 430 BLCA tissues were identified, but the clinical characteristics of only 413 patients could be obtained. We then selected 320 patients with MIBC who either showed recurrence/progression or were disease free. Furthermore, mutational data pertaining to BLCA samples were downloaded from TCGA database.

Data processing

The downloaded BLCA data were normalized using the robust multichip average method for background correction, quantile normalization, and \log_2 transformation [7]. The data were then annotated using the GENCODE (v26) GTF file. On using the DESeq2 R package, we could identify 104 differentially expressed lncRNAs (DElncRNAs) with $|\log_{FC}| > 1$ and adjusted $p < 0.05$ between patients with recurrence and non-recurrence (**Supplementary Fig. S1A**). The samples were then randomly classified into training and validation groups at a 1:1 ratio. The LASSO Cox selection method was used to select the most appropriate number of lncRNAs to categorize the samples into those with recurrence/progression and those that were disease free.

Construction of a recurrence-associated lncRNA signature and nomogram building

Univariable Cox regression survival analysis was used for selecting DElncRNAs associated with survival in bladder cancer. We constructed a recurrence-free survival (DFS) risk-score formula by including eight genes, weighted by their estimated regression coefficients in the univariable Cox regression analysis. Each patient was subjected to the risk-score model and classified into a high- or low-risk group using the median risk score of as the cutoff point. Kaplan–Meier curve and the log-rank test were used for assessing DFS between the high- and low-risk groups. Multivariate analyses with Cox proportional hazards regression were performed on lncRNA risk score, stage, age, invasion depth, grade, gender, and so on. HR and 95% confidence interval (CI) were also calculated. Furthermore, ROC analyses were performed to assess the sensitivity and specificity of recurrence prediction based on the lncRNA risk score, stage, and invasion depth. $p < 0.05$ was considered statistically significant in the log-rank test, Cox regression analysis, and ROC analysis.

Cluster classification using lncRNAs

Using the differential expression profiles of lncRNAs of patients with and without recurrence, bladder cancer subtypes were obtained with unsupervised clustering analysis (the R package ConsensusClusterPlus). After critically evaluating the obtained output (tracking plots, delta plots, and CDF plots), a four-cluster solution was noted to be the most appropriate and informative. We then performed expression clustering analyses using the consensus partitioning around medoids approach, with Pearson correlations, and 10,000 iterations with a 0.95 random fraction of lncRNAs in each iteration. Survival analyses were performed for the four bladder cancer subtypes.

Immunotherapy and chemotherapy response prediction, immune score analysis, immunoinfiltration analysis, and mutational data analysis

Tumor immune dysfunction and exclusion (TIDE) was used to predict immunotherapy response of the four clusters. For computing TIDE scores, the TIDE web application (<http://tide.dfci.harvard.edu/>) was used. The R package “pRRophetic” was applied for predicting chemotherapy response of the four clusters.

The R package ESTIMATE was used to score the bladder cancer samples. The gene expression data of the samples were used to evaluate the stromal and immune scores of each sample through ESTIMATE; further, differences in scores were compared. The R package maxstat was used for determining the optimal cutoff point for continuous variables, and the rank sum test was used to test the significance of

survival between the high- and low-risk groups. The relationship between score and stage, clinical character were demonstrated.

CIBERSORT was used to evaluate the immune infiltrating cells of bladder cancer samples of different subtypes. By inputting the expression data of bladder cancer samples, the proportion of 22 immune infiltrating cells in each sample was obtained, and the sample with $p < 0.05$ was selected for subsequent analysis. We then determined the proportion of the immune infiltrating cells of different subtypes. The relationship between the different immune infiltrating cell ratios and survival within the four clusters and the median DFS of each cluster were assessed.

Mutational data pertaining to bladder cancer samples were downloaded from TCGA database, and the mutation load was calculated using this information. Based on the clinicopathological data, a mutation map of bladder cancer subtypes was constructed to study the types and clinical characteristics of gene mutations. The R package maftools and somatic signature were applied for analyzing the mutations and mapping mutation spectrum and characteristics.

Statistical analysis

Statistical analyses were performed with the R statistical software (R Foundation for Statistical Computing, Vienna, Austria). The primary endpoint for the survival analysis was recurrence-free survival (DFS). DFS was defined as the date of the most recent transurethral resection of the bladder tumor or radical cystectomy till the date of recurrence or death due to any cause. Patients who were lost to follow-up were censored at the date of last contact. We used the Kaplan–Meier method to assess the statistical significance of differences between survival curves for patients in different lncRNA signature groups and molecular subtypes, with the log-rank test.

Results

Acquisition of lncRNA expression datasets

Datasets and correlated clinical data were downloaded from TCGA database. In total, 320 patients with bladder cancer and follow-up data were included (142 patients with recurrence and 178 with non-recurrence). The flow chart was shown in Fig. 1A. We compared the expression data pertaining to bladder cancer with the genome of the human v22 version of the GENCODE database, which led to the identification of 7648 lncRNAs. We then used the R package DESeq2 and applied the significance analysis of microarrays method with $|\log_{2}FC| > 1$ and $p < 0.05$; consequently, from the 7648 lncRNAs, 104 were found to be differentially expressed between patients with and without recurrence. We then randomly categorized the 320 patients with bladder cancer into training ($n = 160$) and validation sets ($n = 160$). The LASSO Cox selection method was applied to select 86 lncRNAs (Supplementary Figure S1B). A ROC curve was constructed to evaluate the classification effect. The area under the ROC curve (AUC) for

the training set was 1 and that for the validation set was 0.792 (Supplementary Figure S1C,D). Therefore, these DElncRNAs were considered to be candidate recurrence-associated lncRNAs.

Development and identification of the lncRNA-based recurrence prediction model and nomogram building

Using the 86 DElncRNAs, we performed univariable Cox regression proportional hazards analysis. Eight lncRNAs related to prognosis were screened out (LINC01449, MGC39584, CTD-2008P7.1, RP5-907D15.4, RP11-789C1.2, RP11-44D19.1, CTC-296K1.4, and AF015262.2; $p < 0.05$). A risk-score formula was derived based on the expression of these eight lncRNAs for DFS prediction: $(0.03589833) \times \text{LINC01449} + (0.002352983) \times \text{MGC39584} + (-3.231363e-05) \times \text{CTD-2008P7.1} + (0.001964683) \times \text{RP5-907D15.4} + (-0.01417278) \times \text{RP11-789C1.2} + (0.02019916) \times \text{RP11-44D19.1} + (-0.000741376) \times \text{CTC-296K1.4} + (0.01122571) \times \text{AF015262.2}$. We then calculated the eight-lncRNA signature risk score for each patient with bladder cancer and ranked them according to their risk scores. The median risk score (0.023) was considered to be the cutoff point to classify the patients into high- ($n = 160$) and low-risk ($n = 160$) groups. The patients in the low-risk group showed significantly longer median DFS than those in the high-risk group (log-rank test $p = 0.0054$; Fig. 1B).

Through a stepwise backward selection process based on AUC, IncScore, stage, and pT remained in the final Cox model for DFS (Fig. 1C). To develop a clinically applicable tool that could provide individualized estimation of 1-, 3-, or 5-year DFS, a nomogram was constructed on the basis of the final Cox model for DFS (Fig. 1D). The nomogram achieved a C-index of 0.719 (95% CI, 0.674–0.764), and the calibration plots showed good consistency between the actual DFS probabilities and the predicted DFS (Fig. 1E,F). Time-dependent ROC analyses also indicated the superior prognostic accuracy of the nomograms (AUC, 0.760; 95% CI, 0.697–0.807; Fig. 1G). These findings validated the importance of the proposed nomograms for DFS prediction.

Consensus classification construction using lncRNAs

Using the differential expression profiles of the 104 lncRNAs, unsupervised consensus clustering was used to derive a robust four-cluster consensus solution (Fig. 2A–C). Survival analysis of the lncRNA-based consensus clusters revealed significant survival differences among the four subtypes (Fig. 2D).

Prediction of immunotherapy and chemotherapy response for the new molecular classification system

TIDE was used to assess the clinical effects of immunotherapy in patients with bladder cancer with the four subtypes; significant differences were found in TIDE values among them ($p = 1.2e^{-14}$; Fig. 3A), which implied that there were differences in the sensitivity of the four subtypes to immunotherapy. A high TIDE

predictive score is associated with a poor effect of immune checkpoint suppression therapy. Cluster 4 showed a higher TIDE predictive score and thus appeared resistant to immunotherapy. In contrast, clusters 2 and 3 showed lower TIDE predictive scores and thus appeared sensitive to immunotherapy. Thus, depending on the subtype, we could determine where immunotherapy would be effective.

“pRRophetic” was used to predict half maximal inhibitory concentration (IC50) values, which reflect the sensitivity of a sample to a drug, of two most commonly used chemotherapy drugs (cisplatin and gemcitabine) for the four subtypes. The sensitivity of the four subtypes to cisplatin (Fig. 3B) and gemcitabine (Fig. 3C) was found to be significantly different. Cluster 4 was sensitive to cisplatin and gemcitabine, while clusters 2 and 3 were resistant to them.

Immune score analysis

The gene expression data of bladder cancer samples were used to evaluate the stromal and immune scores, which represent tumor purity and can be used as indicators of tumor prognosis, using ESTIMATE. A significant difference was found between the scores among the four subtypes (Fig. 4A,B). Cluster 2 showed a high stromal score and was thus associated with low tumor purity and poor prognosis. On the other hand, cluster 4 showed a low stromal score and was accordingly associated with high tumor purity and favorable prognosis. Stromal score also increased with stage, consistent with high score, low purity, and poor prognosis (Fig. 4C). Immune score also increased with an increase in stage (Fig. 4D).

Using maxstat to determine the best cut-off point of continuous variables, the patients were classified into high and low stromal and immune score groups. A significant difference was present between these groups ($p = 0.079$ and 0.037 , respectively; Fig. 4E, F). The patients with a high stromal score showed worse prognosis. Figure 4G depicts a heatmap that reflects relevant data.

Immunoinfiltration analysis

Using the expression (TPM) data, we predicted immune infiltrating cells in different subtypes of bladder cancer with CIBERSORT; consequently, the proportion of 22 immune infiltrating cells in each subtype of bladder cancer was obtained (Fig. 4H). Macrophages M2, T cells CD8, macrophages M0, and macrophages M0 accounted for the largest proportion in clusters 1, 2, 3, and 4, respectively.

To comprehend the relationship between survival and the proportion of immune cells in the four clusters, we performed Cox regression analyses. In cluster 1, mast cells activated were related with survival. According to the median proportion of infiltrating cells, the samples were classified into high- and low-risk groups, and significant differences in survival were found (Supplementary Figure S2A). In cluster 2, mast cells resting and macrophages M1 and M0 were related with survival (Supplementary Figure S2B–D). According to the median proportion of infiltrating cells, the samples were classified into high and low groups, and significant differences in survival were found for only macrophages M0. In cluster 3, T cells

CD4 memory resting, plasma cells, and mast cells resting were related with survival (Supplementary Figure S2E–G). According to the median proportion of infiltrating cells, the samples were classified into high and low groups, and significant differences in survival were found for T cells CD4 memory resting and mast cells resting. Finally, in cluster 4, mast cells activated, macrophages M0, T cells CD4 naïve, and mast cells resting were related with survival (Supplementary Figure S2H–K). According to the median proportion of infiltrating cells, the samples were classified into high and low groups, and significant differences in survival were found for mast cells activated, macrophages M0, and T cells CD4 naïve.

Mutational data analysis

Using the mutational data in TCGA database, we performed statistical analysis using bladder cancer samples, including the type of mutation annotation, proportion of different types of base changes, top 10 mutant genes, and so on (Fig. 5A). Missense mutation is the main type of mutation in bladder cancer; single nucleotide polymorphisms are the main source of mutation, followed by DEL and INS. C > T is the most common single nucleotide polymorphism. The top 10 genes with a high incidence of mutations included, for example, TTN, TP53, and MUC16. The distribution, mutation annotation, survival status (OS, DFS), and tumor mutation load of the four BLCA subtypes (clusters 1–4) are shown in Fig. 5B–E. TTN was the gene with the highest mutation rate in cluster 1, fibroblast growth factor receptor (FGFR)-3 in cluster 2, and TP53 in clusters 3 and 4. As evident from Supplementary Figure S3A–D, the frequency of mutant genes (e.g., TP53 and KDM6A) was different in the four subtypes.

Based on the mutation site of each sample, we considered the bases at the 1-bp position upstream and downstream of the mutation site, divided the mutation into 96 types, and counted the frequency of the 96 mutation types in tumor samples belonging to the four subtypes (Fig. 6A–D).

To determine the relationship between the mutation frequency distribution and signatures collected from COSMIC, we performed non-negative matrix decomposition on the frequency matrix with rows as samples and columns with the 96 mutation types to extract 2 and 3 somatic point mutation characteristics. The similarity analysis of the extracted features and collected signatures revealed that the BLCA mutation spectrum of the bladder cancer subtype 1 (cluster 1) was mainly related to signatures 13, 2, 1, 10, and 5; that for subtype 2 (cluster 2) was mainly related to signatures 10, 2, and 1; that for subtype 3 (cluster 3) was mainly related to signatures 13, 2, and 1; and that for subtype 4 (cluster 4) was mainly related to signatures 13, 2, 6, and 5 (Fig. 6E–H). Signatures 13 and 2 were mainly related to adenosine deaminase, 1 to spontaneous deamination reaction of 5-methylcytosine, 10 to polymerase defects, and 6 to DNA mismatch repair.

Discussion

In this study, we developed a novel tool based on eight lncRNAs to improve the prediction of recurrence in patients with MIBC. The tool could effectively stratify the patients into groups with a low or high risk of recurrence. Furthermore, based on the differential expression of lncRNAs between patients with and

without recurrence, four clusters with different DFS and molecular features were identified. To the best of our knowledge, this is the first study to demonstrate the clinical utility of an lncRNA signature for the postoperative prediction of recurrence in patients with MIBC and to establish an lncRNA-associated cluster with molecular features.

In case of patients with MIBC, RC and pelvic lymph node dissection are the standard methods of treatment [2]. Adjuvant chemotherapy after RC for in these patients with pT3/4 and/or lymph node-positive (N+) disease is still under debate [8]. Local recurrence takes place in soft tissues of the original surgical site or in lymph nodes. Cystectomy is associated with 5–15% probability of pelvic recurrence, which usually occurs during the first 24 months (most often within 6–18 months) after surgery [9]. Distant recurrence is observed in up to 50% patients with MIBC treated using RC. Most importantly, patients generally show poor prognoses post-recurrence [10]. Even with treatment, median survival ranges from 4–8 months after diagnosis. Thus, it is vital to identify high-risk relapse patients and treat them as early as possible. Pertinent risk factors reportedly include pathological stage, positive lymph nodes, positive margins, extent of lymph node dissection, and perioperative chemotherapy without any molecular biological features [11]. However, these factors cannot effectively predict the risk of postoperative recurrence in patients with bladder cancer. Although the TNM staging system and clinical factors are widely used to predict prognoses and guide treatment decisions in such patients, clinical factors are associated with some critical limitations because of the molecular heterogeneity of bladder cancer. Thus, we herein aimed to develop a molecular classification system that used lncScore to categorize samples into subsets with distinct outcomes, facilitating the development of treatment strategies and avoidance of overtreatment or undertreatment. Moreover, we developed a nomogram for individualized assessment of 1-, 3- and 5-year DFS probabilities in patients with MIBC after RC. Our findings therefore suggest that lncScore and recurrence-associated nomogram can be used for predicting the risk of recurrence after RC and for guiding the use of adjuvant chemotherapy to reduce the risk of recurrence.

It has recently been reported that the expression patterns of functional lncRNAs are associated with human cancers [12, 13]. These lncRNAs have been implicated in various tumorigenesis processes, including proliferation, invasion, and apoptosis [14–16]. Some lncRNA-based signatures have been used to predict the risk of cancer progression in patients with renal cell carcinoma and colon cancer [17, 18]. Moreover, prognostic lncRNAs have been previously reported in bladder cancer; four lncRNAs—AC145124.1, AC010168.2, MIR200CHG, and AC098613.1—were reported to form a signature to predict survival in BLCA [19]. Another study reported that a 12-lncRNA-based classifier was related with recurrence in all patients with bladder cancer [20]. However, such studies have been limited by the small number of screened lncRNAs and the lack of prediction of recurrence risk in case of MIBC. They have chiefly focused on both non-muscle invasive bladder cancer and MIBC. In this study, for the first time, we successfully identified an eight-lncRNA signature to predict the risk of recurrence in patients with MIBC. Among the eight lncRNAs, MGC39584 may play a key role in lung squamous cell carcinoma [21]. Furthermore, DNA hypermethylation of the promoter region of GJC1 (CTC-296K1.4), encoding connexin

45, is an important mechanism in silencing gene expression in colorectal cancer [22]. Unfortunately, the other novel lncRNAs have not been investigated in cancer as yet.

Molecular subtyping plays a pivotal role in the study of diseases and for developing personalized therapeutics. The molecular characterization of MIBC by transcriptome profiling has revealed a range of subtypes with distinct clinicopathological characteristics, prognosis, and response to therapeutic regimens. Significant endeavors have been made in the molecular subtyping of MIBC to guide clinical treatment [23–25]. However, previous studies have mainly focused on mRNA-based molecular subtyping; it is noteworthy that mRNA transcripts only represent 1–2% of the transcriptome, which is primarily dominated by ribosomal RNAs and ncRNAs [12]. Thus, we herein selected a list of highly differentially expressed lncRNAs for consensus clustering and identified four clusters with different DFS and molecular features. We found a biologically distinct MIBC subgroup with potential clinical utility. Cluster 1 showed a high frequency of TTN mutation and enriched macrophages M2. Cluster 2 demonstrated potential sensitivity to immunotherapy and insensitivity to chemotherapy, with a high frequency of FGFR-3 mutation and enriched T cells CD8. The high frequency of FGFR-3 mutation implied that cluster 2 may respond to FGFR inhibitors. Novel FGFR inhibitors have been reported to clinically benefit 20% patients with MIBC having tumors harboring mutations or translocations in the tyrosine kinase receptor FGFR-3 and 40% of those having tumors overexpressing FGFR-3. Cluster 3 demonstrated potential sensitivity to immunotherapy and insensitivity to chemotherapy, with a high frequency of TP53 mutation and enriched macrophages M0. Finally, cluster 4 demonstrated potential insensitivity to immunotherapy and sensitivity to chemotherapy, with a high frequency of TP53 mutation and enriched macrophages M0. PD-1–PD-L1 immune checkpoint blockade is becoming standard in patients with locally advanced or metastatic urothelial cancer who relapse after cisplatin-based chemotherapy or are considered cisplatin ineligible (objective response rate, 20%).

This study has some limitations. First, some clinical information was missing (details pertaining to, for example, neoadjuvant and adjuvant treatment). Second, the retrospective nature of this study made it susceptible to inherent biases. Finally, because of the limited sample size, we could not use our lncRNA signature and nomogram to guide clinical treatment. A prospective trial thus needs to be conducted in the future.

Conclusion

To conclude, we report that our eight-lncRNA based signature is a reliable prognostic tool to determine postoperative recurrence in patients with bladder cancer. Moreover, bladder cancer can be stratified into four molecular subtypes that have the potential to guide treatment.

Declarations

Acknowledgements

We would like to thank for the data from TCGA databases.

Author contributions

ZYL, FJZ, and ZWL designed the study. KY, LJJ, ZLZ, MHD, WSW and HCT collected and confirmed the data. SJG and YLY analysed the data. ZYL and LJJ wrote the manuscript. All authors approved the final version of the manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analyzed in the present study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All listed authors have actively participated in the study and have read and approved the submitted manuscript.

Competing Interest

The authors declare that they have no conflict of interest.

Author details

¹State Key Laboratory of Oncology in Southern China, Guangzhou, China. ²Department of Urology, Sun Yat-sen University Cancer Center, No. 651, Dongfeng East Road, Yuexiu District, Guangzhou, Guangdong Province, 510060, China.

³Collaborative Innovation Center for Cancer Medicine, Guangzhou, China.

References

1. Antoni S, Ferlay J, Soerjomataram I, et al. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol*. 2017;71:96–108.
2. Stein JP, Lieskovsky G, Cote R, 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.
3. Spizzo R, Almeida MI, Colombatti A, et al. Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene*. 2012;31:4577–87.
4. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*. 2009;136:629–41.
5. Hauptman N, Glavac D. Long non-coding RNA in cancer. *Int J Mol Sci*. 2013;14:4655–69.
6. Qi D, Li J, Que B, et al. Long non-coding RNA DBCCR1-003 regulate the expression of DBCCR1 via DNMT1 in bladder cancer. *Cancer Cell Int*. 2016;16:81.
7. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249–64.
8. Cohen SM, Goel A, Phillips J, et al. The role of perioperative chemotherapy in the treatment of urothelial cancer. *Oncologist*. 2006;11:630–40.
9. Soukup V, Babjuk M, Bellmunt J, et al. Follow-up after surgical treatment of bladder cancer: a critical analysis of the literature. *Eur Urol*. 2012;62:290–302.
10. Huguet J. Follow-up after radical cystectomy based on patterns of tumour recurrence and its risk factors. *Actas Urol Esp*. 2013;37:376–82.
11. Ghoneim MA, Latif MA, et al. Radical cystectomy for carcinoma of the bladder: 2,720 consecutive cases 5 years later. *J Urol*. 2008;180:121–7.
12. Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer*. 2011;10:38.
13. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1:391–407.
14. Bo C, Li N, Li X, et al. Long noncoding RNA uc.338 promotes cell proliferation through association with BMI1 in hepatocellular carcinoma. *Hum Cell*. 2016;9:41–7.
15. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464:1071–6.
16. Khaitan D, Dinger ME, Mazar J, et al. The melanoma-upregulated long noncoding RNA SPRY4-IT1 modulates apoptosis and invasion. *Cancer Res*. 2011;71:3852–62.
17. Qu L, Wang ZL, Chen Q, et al. Prognostic Value of a Long Non-coding RNA Signature in Localized Clear Cell Renal Cell Carcinoma. *Eur Urol*. 2018;74:756–63.
18. Dai W, Feng Y, Mo S, et al. Transcriptome profiling reveals an integrated mRNA-lncRNA signature with predictive value of early relapse in colon cancer. *Carcinogenesis*. 2018;39:1235–44.

19. He RQ, Huang ZG, Li TY, et al. RNA-Sequencing Data Reveal a Prognostic FourlncRNA-Based Risk Score for Bladder Urothelial Carcinoma: An in Silico Update. *Cell Physiol Biochem*. 2018;50:1474–95.
20. He A, He S, Peng D, et al. Prognostic value of long non-coding RNA signatures in bladder cancer. *Aging*. 2019;11:6237–51.
21. Shi YF, Li YP, Yan C, et al. Identification of Key Genes and Evaluation of Clinical Outcomes in Lung Squamous Cell Carcinoma Using Integrated Bioinformatics Analysis. *Oncol Lett*. 2019;18:5859–70.
22. Sirnes S, Honne H, Ahmed D, et al. DNA Methylation Analyses of the Connexin Gene Family Reveal Silencing of GJC1 (Connexin45) by Promoter Hypermethylation in Colorectal Cancer. *Epigenetics*. 2011;6:602–9.
23. Robertson AG, Kim J, Al-Ahmadie H, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell*. 2017;171:540–56.
24. Hurst CD, Knowles MA. Bladder cancer: multi-omic profiling refines the molecular view. *Nat Rev Clin Oncol*. 2018;15:203–4.
25. Choi W, Porten S, Kim S, et al. Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell*. 2014;25:152–65.

Figures

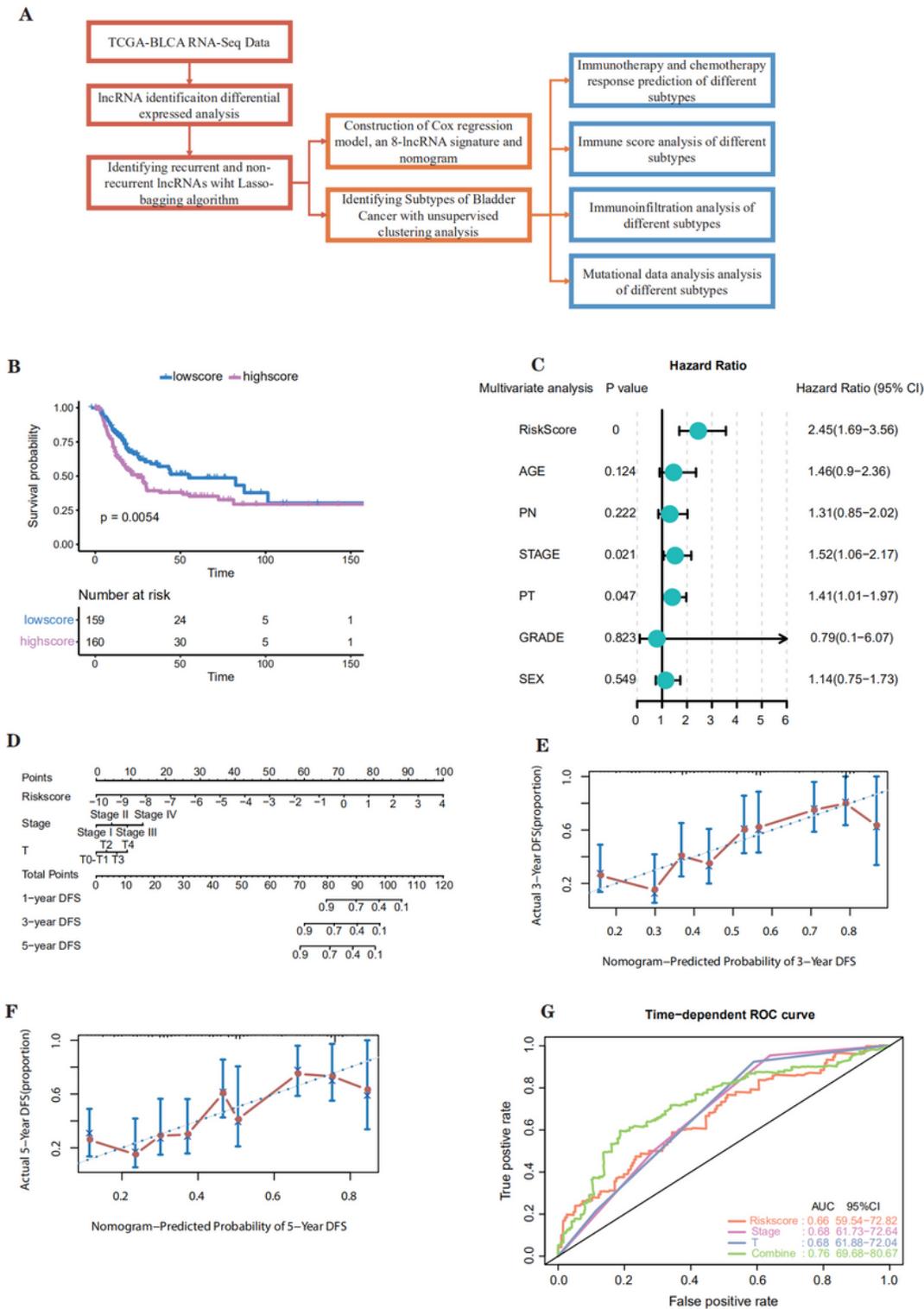


Figure 1

1A Flowchart of Study; Figure 1B Kaplan–Meier curves of DFS based on the lncScore in MIBC patients; Figure 1C Construction of a Cox model for DFS; Figure 1D Nomogram to predict the DFS of MIBC patients; Figure 1E Calibration curves of the nomogram to predict the 3-year DFS; Figure 1F Calibration curves of the nomogram to predict the 5-year DFS; Figure 1G Prediction of DFS by Time-dependent ROC analyses

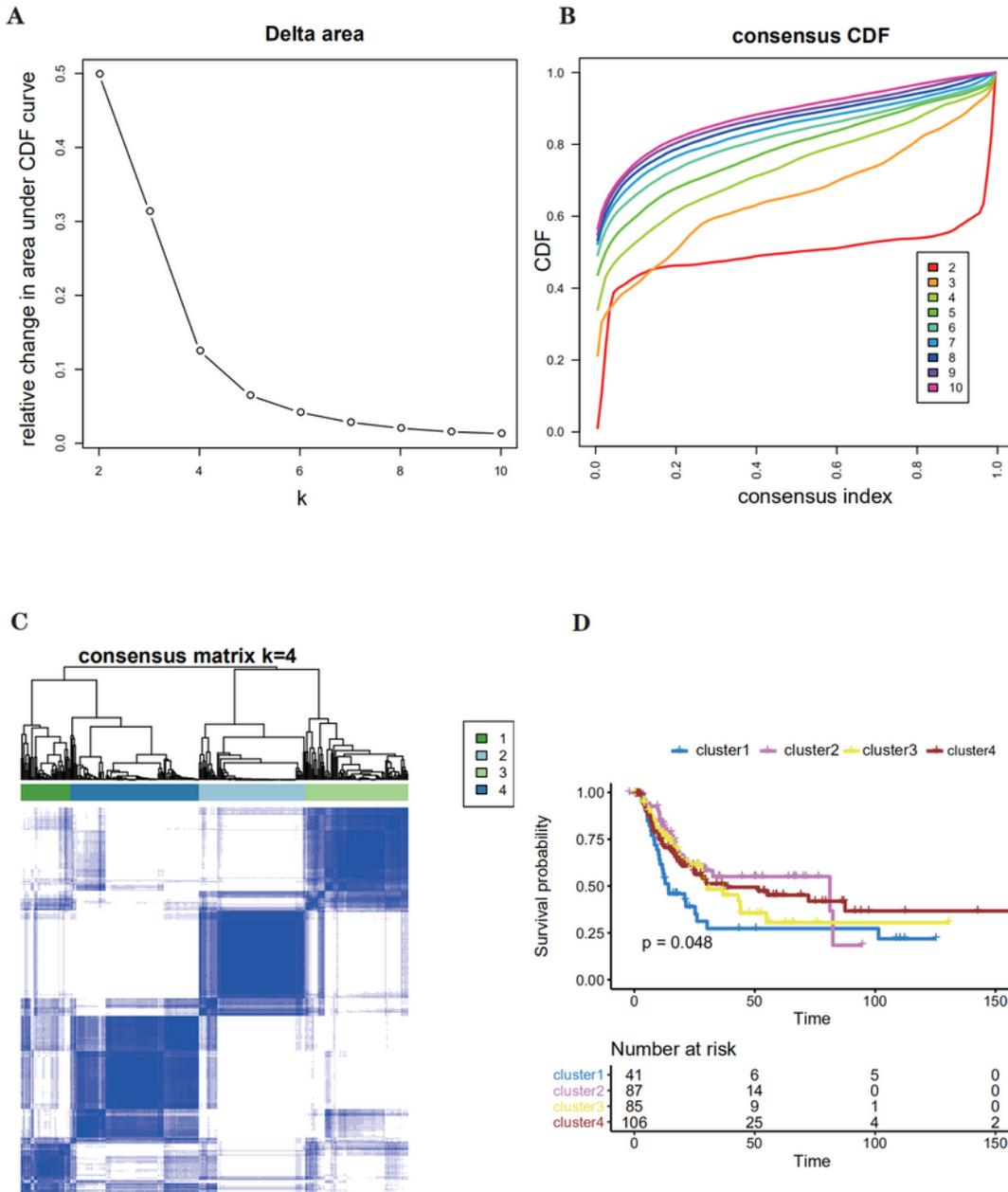


Figure 2

2A Optimal k-value selection graph for consistent clustering; Figure 2B Consistent clustering CDF graph; Figure 2C Sample clustering heat map for k = 4; Figure 2D Kaplan–Meier curves among the four subtypes of bladder cancer

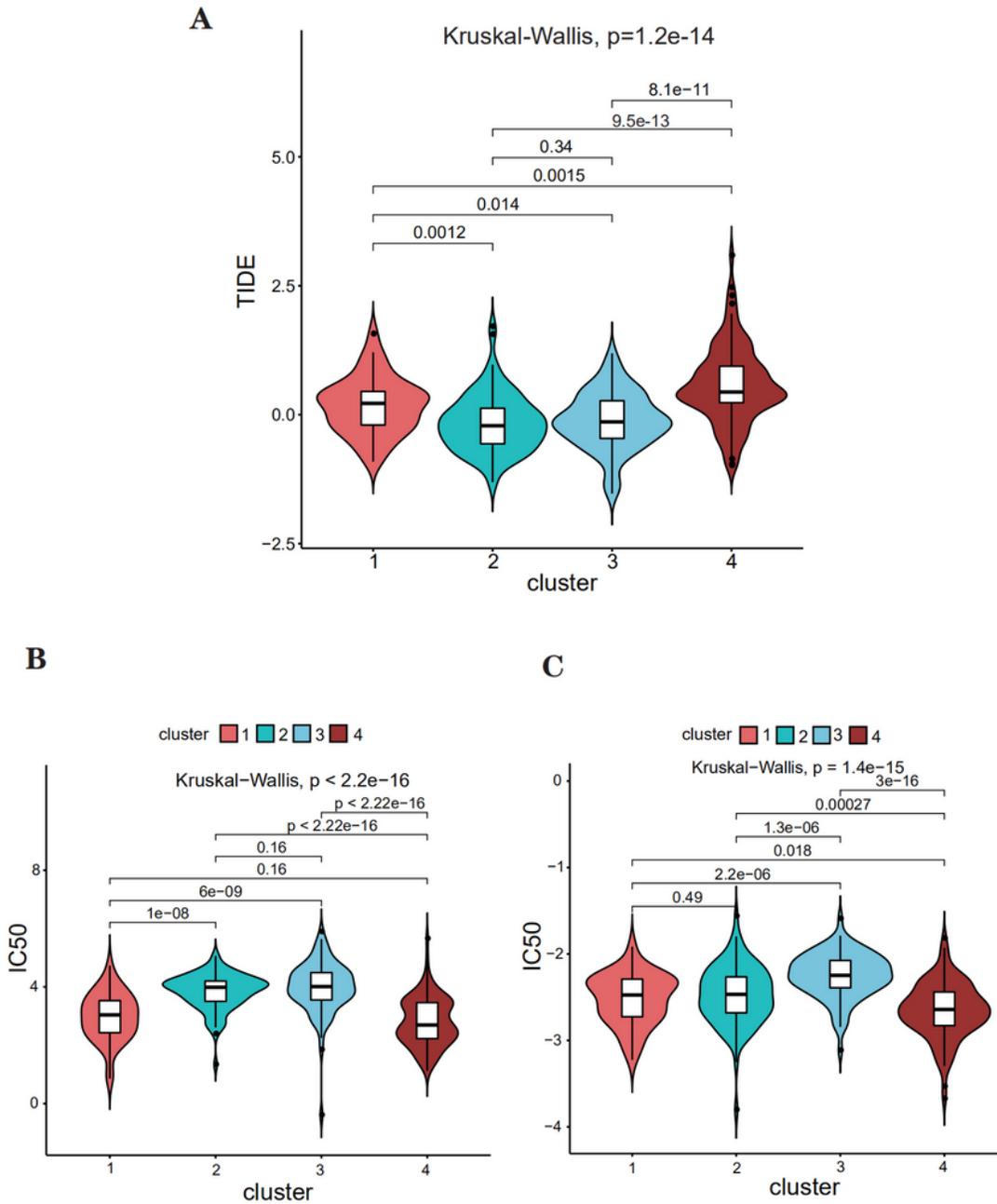


Figure 3

3A Prediction of ICI therapy for four subtypes of bladder cancer Figure 3B Prediction of cisplatin for four subtypes of bladder cancer Figure 3C Prediction of gemcitabine for four subtypes of bladder cancer

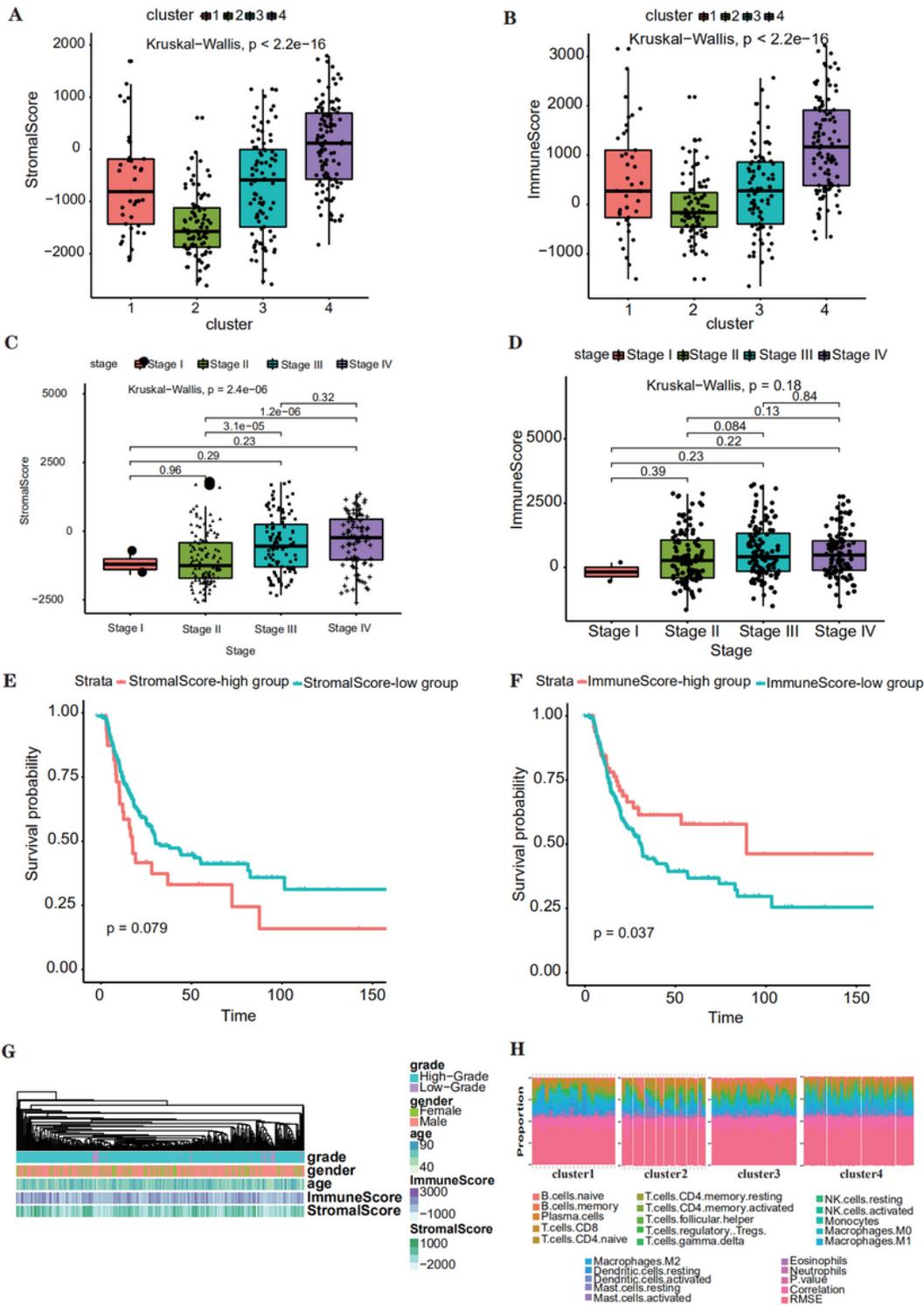


Figure 4

4A StromalScore of four subtypes of bladder cancer; Figure 4B ImmuneScore of four subtypes of bladder cancer; Figure 4C StromalScore of different stages of bladder cancer; Figure 4D ImmuneScore of different stages of bladder cancer; Figure 4E Kaplan–Meier curves among high and low StromalScore group; Figure 4F Kaplan–Meier curves among high and low ImmuneScore group; Figure 4G Heatmap of grade,

gender, age and StromalScore, ImmuneScore; Figure 4H Proportions of immune infiltrating cells among the four subtypes of bladder cancer

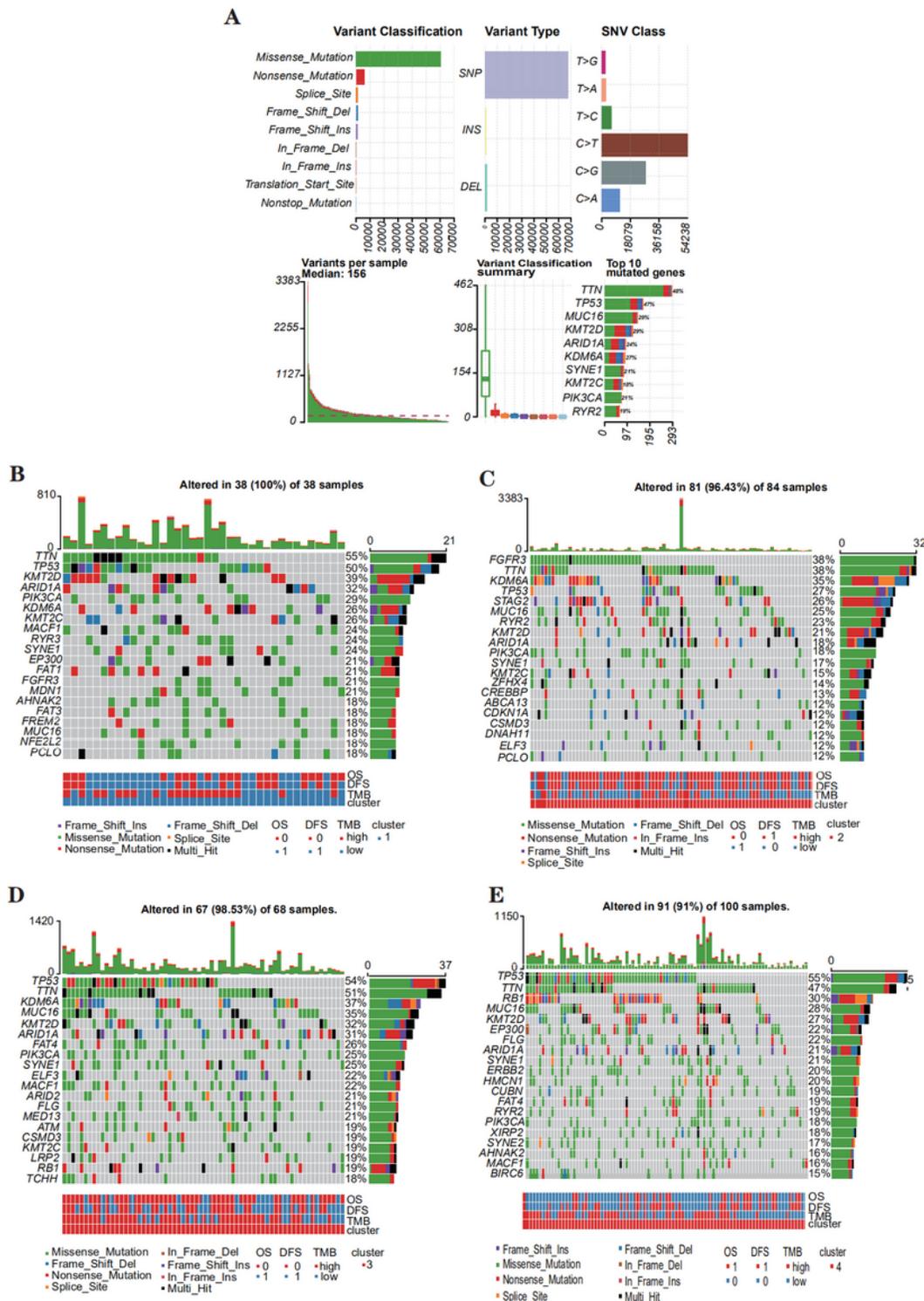


Figure 5

5A Bladder cancer mutation profile; Figure 5B Distribution of common gene mutations in cluster 1 of bladder cancer; Figure 5C Distribution of common gene mutations in cluster 2 of bladder cancer; Figure

5D Distribution of common gene mutations in cluster 3 of bladder cancer; Figure 5E Distribution of common gene mutations in cluster 4 of bladder cancer

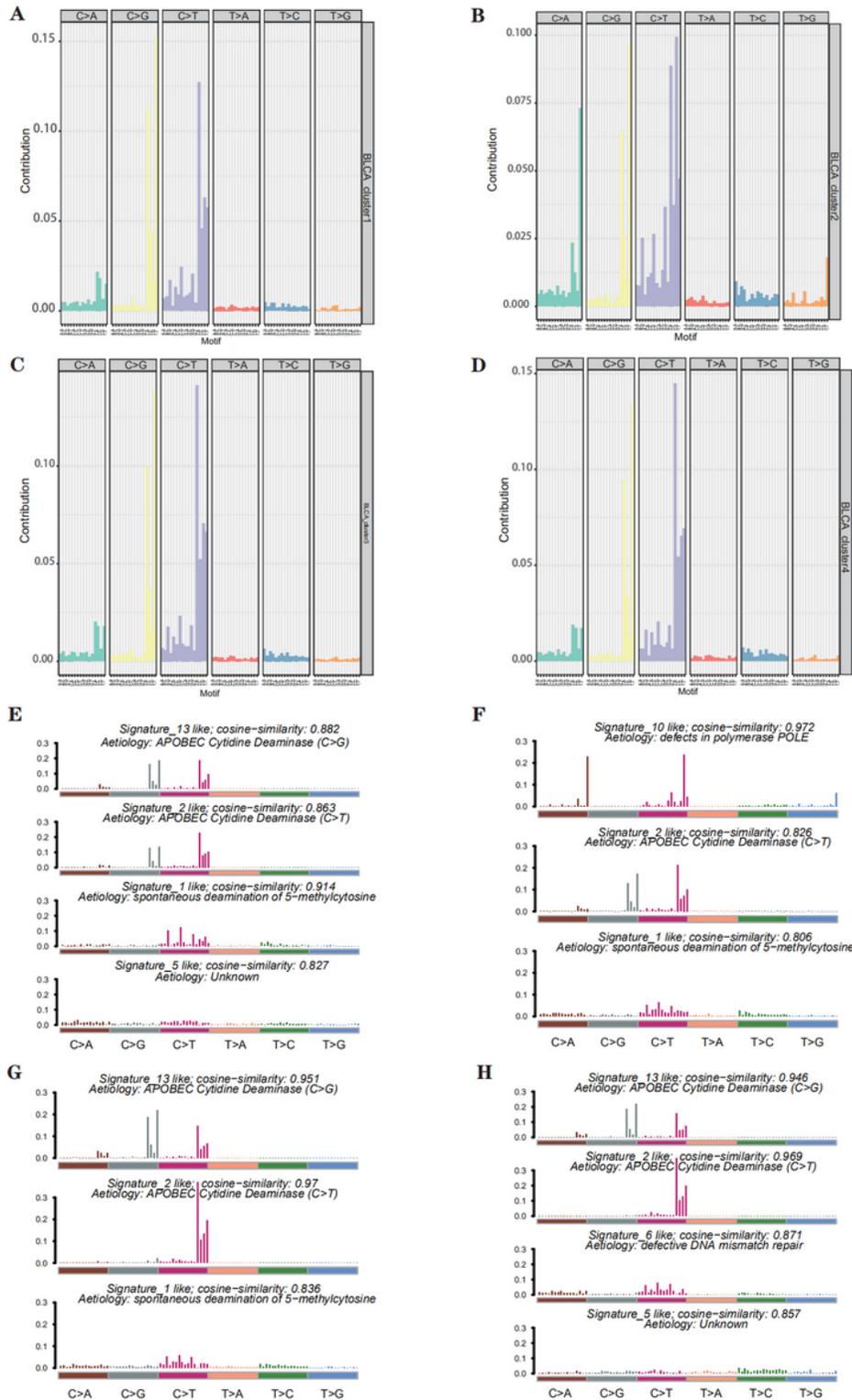


Figure 6

6A Similarity between mutation characteristics of bladder cancer cluster 1 and cosmic mutation signature; Figure 6B Similarity between mutation characteristics of bladder cancer cluster 2 and cosmic mutation signature; Figure 6C Similarity between mutation characteristics of bladder cancer cluster 3 and

cosmic mutation signature; Figure 6D Similarity between mutation characteristics of bladder cancer cluster 4 and cosmic mutation signature; Figure 6E Frequency distribution of 96 mutation types in cluster 1 of bladder cancer; Figure 6F Frequency distribution of 96 mutation types in cluster 2 of bladder cancer; Figure 6G Frequency distribution of 96 mutation types in cluster 3 of bladder cancer; Figure 6H Frequency distribution of 96 mutation types in cluster 4 of bladder cancer

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1.png](#)
- [S2.png](#)
- [S3.png](#)