

DNA methylation-based classification and identification of bladder cancer prognosis-associated subgroups

Zijian Tian

Beijing Hospital

Lingfeng Meng

Beijing Hospital

Xingbo Long

Beijing Hospital

Tongxiang Diao

Beijing Hospital

Maolin Hu

Beijing Hospital

Miao Wang

Beijing Hospital

Ming Liu (✉ liumingbjyy@126.com)

Beijing Hospital <https://orcid.org/0000-0001-8139-8804>

Jianye Wang (✉ wangjybjyy@126.com)

Beijing Hospital

Primary research

Keywords: DNA Methylation, urinary bladder neoplasms, epigenetics, cluster analysis

Posted Date: April 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-19933/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published on June 17th, 2020. See the published version at <https://doi.org/10.1186/s12935-020-01345-1>.

Abstract

Background : Bladder cancer (BCA) is the most common urinary tumor, but its pathogenesis is unclear, and the associated treatment strategy has rarely been updated. In recent years, a deeper understanding of tumor epigenetics has been gained, providing new opportunities for cancer detection and treatment.

Methods : We identified prognostic methylation sites based on DNA methylation profiles of BCA in the TCGA database and constructed a specific prognostic subgroup.

Results : Based on the consistent clustering of 402 CpGs, we identified seven subgroups that had a significant association with survival. The difference in DNA methylation levels was related to T stage, N stage, M stage, grade, sex, age, stage and prognosis. Finally, the prediction model was constructed using a Cox regression model and verified using the test dataset; the prognosis was consistent with that of the training set.

Conclusions : The classification based on DNA methylation is closely related to the clinicopathological characteristics of BCA and determines the prognostic value of each epigenetic subtype. Therefore, our findings provide a basis for the development of DNA methylation subtype-specific therapeutic strategies for human bladder cancer.

1. Background

Bladder cancer (BCA) is the ninth most common cancer in the world [1], and it is estimated that the number of new BCA cases in the United States will reach 81,400 in 2020, with approximately 17,980 deaths [2]. Notably, the clinical course of BCA varies with the degree of tumor invasion. In patients with BCA, 70% of cases are non-muscular invasive BCA (NMIBC), characterized by a high recurrence rate and low mortality. The remaining 30% are muscular invasive BCA (MIBC), which is prone to early metastasis, and approximately half of these cases are fatal [3, 4]. Cystoscopy is the current gold standard to monitor BCA. However, this method involves a highly invasive examination and has a sensitivity of 85% for the diagnosis of exophytic tumors [5]. It is worth noting that the postoperative pathology of transurethral bladder tumors does not always accurately reflect the tumor stage, due to limitations such as the experience of the operator and destruction of tissue caused by energy instruments. Studies have shown that the proportion of transurethral surgical procedures that underestimate tumor stage could be as high as 25% [6, 7]. Further, repeated cystoscopy is used for continuous follow-up, which makes BCA one of the most expensive malignant tumors to treat [5, 8]. At the same time, in the absence of a new targeted therapy-based strategy, despite surgery and cisplatin-based chemotherapy, it is clear that the existing treatment modality has reached its peak, with only a slight improvement in patient survival but more side effects [9]. As the survival time of patients with NMIBC progression to MIBC is poor [10], it is very important to better understand the biological pathogenesis of BCA, to more accurately monitor its progression, predict patient prognosis, and choose appropriate treatments.

Therefore, the existing diagnosis, prognosis, and monitoring systems for BCA patients need to be improved. Epigenetics comprise heritable but reversible modifications that can alter gene expression without changing the original DNA sequence. The maintenance of epigenome function is the basis of normal gene expression. Changes in epigenome function will affect the basic processes of cell proliferation, differentiation, and death, which might lead to cancer [11]. Therefore, the diagnosis, the evaluation of prognosis and the prediction of response to treatment can prospectively count on cancer biomarkers based on epigenetics. [12]. DNA methylation was the first epigenetic modification found in humans in the early 1980s [13]. Among the epigenetic mechanisms, DNA methylation is the best studied, and abnormal CpG island methylation was proven to be related to the occurrence and development of many cancer types, including BCA [14]. At the same time, several studies also support this argument. Specifically, 90% sensitivity and 93% specificity were observed when evaluating the methylation of Twist1 and NID2 [15]. The methylation status of SOX-1, IRAK3, and LI-MET genes also showed better ability to predict progression than cystoscopy [16]. Uromark is described as a novel next-generation sequencing-based biomarker, based on 150 CpGs, with a sensitivity of 98% and a specificity of 97% for monitoring BCA [17]. Further, in terms of prognosis, hypermethylated TIG1, GSTP1 and APC in BCA patients were found to be associated with poor survival outcomes, showing 93% specificity and 80% sensitivity [18].

However, it is recognized that the specific methylated sequence of the gene promoter region has not been identified yet. Therefore, the objective of this study was to identify DNA methylation profiles in BCA from the TCGA database and to identify biologically and clinically relevant molecular subsets. Our classification scheme could help to identify new BCA molecular subtypes and prognostic model based on methylation site to accurately subdivide BCA patients and improve clinical prognostic assessments and personalized treatment.

2. Methods

2.1 Data selection and pre-processing

For this study, Samples from the TCGA database containing 437 BCA methylation data are downloaded from UCSC Cancer Browser (<http://xena.ucsc.edu/>, 2020-02-23). RNA-sequencing data from 433 BCA samples were downloaded from TCGA (<https://cancergenome.nih.gov/>, 2020-02-23), and among these, 407 samples were associated with clinical data. CpGs with missing data in more than 70% of the samples were excluded from the analysis. Based on the polymorphic CpGs and cross-reaction probe, the CpGs of the cross-reaction probe in the genome was also removed. The k-nearest neighbor imputation method in SVA R software package was used to estimate other unrecognized probes [19]. We also removed unstable genomic sites with CpGs and single nucleotide polymorphisms in sex chromosomes. We only studied the CpGs in the promoter region because DNA methylation in the promoter region (2 kb upstream of the transcriptional initiation site to 0.5 kb downstream) significantly affects gene expression. Finally, the BCA samples were randomly grouped as 203 training samples and 204 testing samples.

2.2 Using Cox proportional hazard regression model to determine CpGs of prognosis

First, a univariate Cox proportional hazard regression model was established based on CpGs, T stage, N stage, M stage, grade, age, sex, stage, and survival data ($P < 0.001$). The significant CpGs obtained from the univariate Cox proportional hazard regression model were used to analyze the multivariate Cox proportional hazard regression mode ($P < 0.001$). Finally, CpGs that were significantly modulated in multivariate Cox regression analyses were selected as characteristic CpGs.

2.3 Establishment of molecular subtypes using consensus clustering

Based on the most variable CpG sites, the K-means clustering algorithm in ConsensusClusterPlus R packet [20] was used to perform consistent clustering to identification BCA subgroup. The pre-specified dataset was classified into k clusters via the algorithm. Here, 80% of the tumor samples were selected in each iteration, and the K-means algorithm and Euclidean square distance measure served as the grouping of Kappa 2 and 20. We used more than 100 iterations to determine the stability of each cluster. At the same time, the maximum number of clusters which have at least 90% cluster consensus was selected.

The optimal number of clusters was defined based on the delta area map and the cumulative distribution function. Color gradients were used to describe common values from 0 to 1 (white to dark blue). Items belonging to the same cluster in the matrix should be adjacent to each other. Therefore, the diagonal blue blocks on the white background in the color-coded heat map correspond to the matrix of perfect consensus. The best cluster number need to meet the following requirements: high consistency, small coefficient of variation, and the area under the cumulative distribution function (CDF) curve not increasing remarkably. The number of categories was defined by the relatively insignificant change in the area under the CDF curve. The heat map corresponding to consensus clustering was generated with the pheatmap R packet.

2.4 Survival and clinical characteristics analyses

Construction of the overall survival (OS) curve of BCA subsets defined by the DNA methylation profile was performed via the Kaplan-Meier method. Significant differences among clusters were evaluated by the log-rank test. A comprehensive analysis of the relationship between clinical information and biological characteristics of clusters based on DNA methylation was performed through a chi-square test. R Bioconductor survival package was used for survival analysis. $P \leq 0.05$ was considered statistically significant.

2.5 Gene Ontology enrichment and KEGG pathway analysis

We use the clusterProfiler software package in R [21] for Gene Ontology (GO) enrichment and KEGG analysis to obtain GO terms (biological processes, molecular function, and cellular component) and KEGG pathways. Significance was defined as $P < 0.05$.

2.6 Establishment and verification of the prognostic prediction model

According to the `coxph` function of the survival package in R, combined with the methylation characteristics and prognostic information of the three CpGs, a Cox proportional hazard model was constructed. For this model, we use the formula as: $\text{risk score} = 2.4139 \times \text{cg18312429} + 2.0075 \times \text{cg19826026} + 2.4836 \times \text{cg27562023}$. The ROC packet in R was used to establish a receiver operation characteristic (ROC) curve. To verify the reliability as well as stability of this model, 204 testing samples were analyzed using this prediction model.

2.7 Generation and validation of a predictive nomogram

In this study, univariate and multivariate Cox regression analyses were performed to determine independent prognostic factors to construct the nomogram. All independent prognostic factors obtained by multivariate COX regression analysis were selected to construct a combined prognostic model to evaluate the probability of 1-year, 3-year, and 5-year OS in patients with BCA. To evaluate the consistency between the actual survival rate and the predicted survival rate, a calibration curve can be drawn in which the 45° line represents the best prediction and the coincidence represents the better predicted value of the model.

3. Results

3.1 Identification prognostic methylation sites associated with OS

The prognostic molecular subsets were clustered based on a DNA methylation map of BCA samples in the TCGA database. Successfully, a total of 21,122 CpG sites were screened for methylation based on 203 experimental samples in the training group. The univariate Cox proportional hazard regression model was applied to each CpG site in the training group. Here, 1096 important CpG sites that might affect the survival of patients ($P < 0.001$) were identified. Accordingly, 1096 significant CpG sites were taken into the multivariate Cox proportional hazard regression model to study independent prognostic factors ($P < 0.001$). Finally, 402 meaningful CpG sites were obtained for further prognostic subgroup analysis (Supplementary Table 1).

3.2 Identification of different DNA methylation prognostic subgroups through consensus clustering and prognostic analyses between clusters

The consensus clustering of 402 potential prognostic methylation sites was performed to identify different DNA methylation molecular subsets and for further prognostic analyses. The optimal class number was obtained by calculating the average cluster consistency and intercluster coefficient of variation of each class number. As displayed by the CDF curve, when the clustering number is 7, the clustering result is relatively stable (Figure 1a). Further observation of the CDF delta area curve shows that the area under the CDF curve starts to remain stable after seven categories, suggesting that when the cluster is selected as 7, the cluster has stable clustering results, and the seven categories were also considered the ideal number of categories for further analysis (Figure 1b). The consensus matrix, as shown in Figure 2a, illustrates the consensus of $K = 7$ and shows the superlative seven-block structure.

Moreover, the heatmaps in accordance with 402 CpGs were generated via the heatmap function based on DNA methylation classification with T category, N category, M category, grade, stage, sex and age (Figure 2b). The Kaplan–Meier diagram illustrated that the prognosis of BCA, defined by consensus clustering based on methylation, had significant differences among the seven clusters ($P < 0.05$; Figure 3a). According to our results, the prognosis of Clusters 2 and 6 were the best, whereas that of Cluster 7 was the worst. Figure 3b-h indicates the intracluster proportions for the seven clusters on the basis of T category, N category, M category, grade, stage, sex, and age. The association trends between features and specific clusters were as follows: Clusters 1 and 7 were in advanced stage; Clusters 2 had lower T grade; Clusters 7 had higher N grade, younger age, and more male; Cluster 5 was in higher M grade; Clusters 1, 5, and 7 had higher grade.

3.3 Differential feature recognition based on DNA methylation clustering and screening of cluster-specific methylation sites

A total of 557 promoter genes were identified according to the genome annotation of 402 meaningful CpG sites. The most significant GO terms for biological process, cellular component, and molecular function are shown in Figure 4a-c. Next, we analyzed the functional enrichment of these 557 genes and recognized 23 significantly enriched pathways ($P < 0.05$; Figure 4d). We observed that the three most significantly enriched pathways were peroxisome, base excision repair, and fatty acid biosynthesis pathways.

The expression of methylated genes found in the subgroup were also studied. Expression values were available for 465 of the 557 genes in the training group. We also generated a heatmap for specific annotated genes with methylation sites, as shown in Figure 5, with different gene expression patterns among different subgroups, indicating that the level of DNA methylation usually reflects the expression of these genes.

The cluster-specific methylation sites were screened by using methylation sites as the feature of the clusters. The differences among seven clusters were analyzed for each methylation site. Eighteen cluster-specific methylation sites were identified and are shown in the heatmap in Figure 6. Clusters 1, 2, and 6 had more specific sites, Cluster 1 had the highest level of methylation among all clusters, and Cluster 2 had the lowest methylation level among all clusters (Figure 7).

3.4 Construction and validation of the prognosis prediction model

Cluster 2 was found to contain a great number of samples that were associated with good prognosis and the highest number of specific methylation sites, which made it as seed cluster. Using the formula provided in the Methods section, we constructed a Cox proportional hazard model based on the methylation distribution of three specific sites combined with prognostic information. As shown in Figure 8a, there was a significant difference in prognosis between the two groups. The results of ROC analysis using the risk score calculated for each sample are shown in Figure 8b. The area under the curve (AUC) of the model was 0.788, which was higher than that of other predictive factors, indicates that the function of

the model is effective. The samples were divided into high-risk group and low-risk group by median risk score which was used as cutoff value and then generate the risk curve (Figure 8c).

Next, the prognostic model was used to predict the outcomes of patients in the test dataset. The methylation level curves of three CpGs were obtained to test the dataset samples, and the prognostic model was used to calculate the risk score. The test dataset samples were then drawn into high-risk group and low-risk group. It is worth noting that there was a significant difference in prognosis between the two groups ($P = 0.00666$). It was consistent with the results gained from the training data set, indicating the accuracy of prediction and stability of the model (Figure 9a-c).

3.5 Clinical application of a nomogram

Through univariate and multivariate Cox regression analyses (Figure 10a-b), both our prognostic model and sex were identified as potentially independent predictors, based on which nomogram was constructed (Figure 11a). The C index was 0.77, and the correction chart showed that the actual survival rate of 5 years was in good agreement with the predicted survival rate (Figure 11b). This indicates that the nomogram has high potential for clinical application.

Discussion

The classical view of cancer evolution is that a series of genetic changes promotes the transition from the early precancerous stage to invasive cancer and affects the incidence of metastatic diseases. During carcinogenesis, oncogenes can be activated to promote cell division or inhibit cell death. At the same time, tumor suppressor genes can be inactivated in a way that can promote abnormal cell proliferation. Therefore, both the functional gain of proto-oncogene mutations and loss-of-function mutations in tumor suppressor genes might cause cancer through uncontrolled cell growth and defective apoptosis [22]. DNA methylation involves the addition of a methyl group at the cytosine 5' carbon position of CpG dinucleotides in the genome, which is an important element of epigenetic regulation of gene expression [23]. Since the 1990s, increasing number of studies have recognized that heritable changes regulated by epigenetics might also play a vital role in the evolution of all types of human cancer [24]. We now know that epigenetic changes occur through specific events, including early widespread loss of normal DNA methylation and an increased number of focal gains in gene promoters [11].

BCA is a malignant tumor of the bladder that originates from the transformation of transitional intraepithelial urothelial cells, and is therefore also known as urothelial carcinoma or transitional cell carcinoma. Although transitional cell carcinoma of the bladder ranks fourth among men [2], the mechanism of the occurrence and development of urothelial carcinoma is still not completely clear. The development of modern techniques for genome-wide DNA methylation detection enables a more in-depth analysis of BCA methylation. Wolff et al. confirmed that the majority of DNA methylation changes occurred in the early stage of BCA which were conserved in carcinoma in situ, non-invasive tumors and invasive tumors, and were located on the CpG island [25]. Compared to the urothelium from a healthy bladder, the hypermethylation of *ZO2*, *MYOD*, and *CDH13* were also detected in the urothelium with a

normal appearance in patients with BCA, suggesting that epigenetic 'field defects' might be one of the reasons for the loss of epithelial integrity. Changes in DNA methylation comprise an early driver of cancer, and epigenetic changes involving DNA methylation might result in subsequent genome changes, which create a permissible environment for the onset and recurrence of BCA [25, 26]. In an interesting study, the gene methylation pattern of secondary bladder recurrence of primary upper urinary tract cancer was tested and it was confirmed that the methylation rate of some genes increased with the increase in the number of recurrences, which might be a predictor of postoperative recurrence [27]. Further, the methylation status of *GP5* and *ZSCAN12* can effectively be used to distinguish between high-grade and low-grade BCA [28]. At the same time, the methylation level of genes can effectively identify the degree of invasion of bladder tumors [25]. It is worth noting that based on TCGA data, the level of methylation and expression of *SOWAHC* is associated with prognosis [29]. *HOXA9* promoter methylation has also been shown to be associated with an increase in recurrence and progression in NMIBC. Importantly, it was also proved to be related to cisplatin resistance in BCA cells [30, 31]. In distinguishing whether patients with BCA have lymph node metastasis, a three-gene methylation panel was shown to predict the progression of metastasis and allow patients to benefit from lymphadenectomy and neoadjuvant chemotherapy [32].

We used CpG sites to identify seven different prognostic subtypes of BCA, which could predict survival of the disease, as well as the TNM classification, grade, stage, and age distribution of prognosis among the seven molecular subtypes. Thus, this classification method results in molecular stratification that is suitable for a single tumor, which has an important impact on treatment decisions and accurate diagnoses. According to the seven molecular subtypes, if the important CpG sites were classified into category 7, the patients were found to have poor staging (more inclined to Stage IV), higher grade, a higher probability of lymph node metastasis, and poor prognosis. This is of great significance for early intervention and to actively encourage patients to receive treatment. Therefore, the radical resection of BCA and early lymph node dissection can be performed more actively. If the sequence of the CpG sites are classified into category 5, the risk of tumor metastasis is higher. Therefore, radiotherapy and chemotherapy should be actively performed to reduce the incidence of metastasis. However, when the sequence of the CpGs are classified as category 2, the tumor has lower invasiveness and the patient has a higher 5-year survival rate, indicating a better prognosis. This classification can motivate doctors to reconsider the individualized treatment of patients, conduct close clinical follow-up, and minimize overtreatment, which would help to reduce pain for the patients. In summary, our study of these seven subtypes at the DNA methylation molecular level indicated that this system could be used to more accurately classify BCA and guide clinicians in the diagnosis, treatment, and prognosis of different epigenetic subtypes.

With the accumulation of knowledges on the biology, function, and regulatory mechanisms of epigenetic modifications in cancer, and considering the limited efficacy of chemotherapy and immunotherapy, clinical success has ushered in an era of epigenetic therapy aimed at reactivating genes that are improperly silenced during carcinogenesis [33]. One of the first identified histone lysine methyltransferase inhibitors specific for G9a (EHMT2) is BIX-01294, which was considered to inhibit the proliferation of BCA cell lines [34]. In recent years, CM272 has been considered a novel double inhibitor of G9a/DNMT1, and it

has significant antitumor effects on BCA *in vivo* and *in vitro* [35]. Demethylation drugs can inhibit the proliferation, migration, and invasion of BCA cells and have been shown to enhance the sensitivity of cisplatin-resistant cells [31, 36]. Using these concepts, DNA demethylation reagents are expected to be used in new therapeutic applications and might play an important role in cancer treatment. Therefore, the cancer epigenome as a target for cancer treatment still represents a vital opportunity for clinical exploitation, and exciting progress is expected in the next few years.

Conclusions

We identified the methylation sites related to prognosis in patients with BCA and constructed a predictive model for these patients. The molecular subtypes based on methylation sites were found to be closely related to clinicopathology and could better predict the progression of BCA. Finally, this new targeting strategies will bring new horizons for BCA management and detection.

Abbreviations

AUC: area under the curve, BCA: bladder cancer, CpGs: CpGs, MIBC: muscular invasive bladder cancer, NMIBC: non-muscular invasive bladder cancer, ROC: receiver operation characteristic

Declarations

Authors' contributions

Study concept & design: Zijian Tian and Jianye Wang; data acquisition or data analysis/interpretation: Lingfeng Meng and Ming Liu; manuscript drafting or manuscript revision for important intellectual content: all authors; approval of final version of submitted manuscript: all authors; literature research: Xingbo Long, Tongxiang Diao and Maolin Hu; manuscript editing: all authors.

Acknowledgments

The results published here are based on data generated by the TCGA Research Network (<http://cancergenome.nih.gov/>) and GEO Research Network (<http://www.ncbi.nlm.nih.gov/geo/>). We would like to thank Editage (www.editage.cn) for English language editing.

Funding sources

This study was financially supported by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (grant number: 2018-I2M-1-002) and the Beijing Hospital Clinical Research 121 Project (BJ-2018-090). The funders had no roles in study design, data collection, data analysis and interpretation, or writing the manuscript.

Competing interests

The authors have declared no competing interests.

Ethics Approval and consent to participate

Not applicable

Consent for Publication

Not applicable

Availability of Data and Materials

For this study, Samples from the TCGA database containing 437 BCA methylation data are downloaded from UCSC Cancer Browser (<http://xena.ucsc.edu/>, 2020-02-23). RNA-sequencing data from 433 BCA samples were downloaded from TCGA (<https://cancergenome.nih.gov/>, 2020-02-23).

References

1. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder cancer incidence and mortality: A global overview and recent trends. *Eur Urol.* 2017;71:96-108.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70:7-30.
3. Knowles MA, Hurst CD. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer.* 2015;15:25-41.
4. Alfred Witjes J, Lebet T, Comperat EM, Cowan NC, De Santis M, Bruins HM, et al. Updated 2016 EAU guidelines on muscle-invasive and metastatic bladder cancer. *Eur Urol.* 2017;71:462-75.
5. Lodewijk I, Duenas M, Rubio C, Munera-Maravilla E, Segovia C, Bernardini A, et al. Liquid biopsy biomarkers in bladder cancer: A current need for patient diagnosis and monitoring. *Int J Mol Sci.* 2018;19.
6. Ark JT, Keegan KA, Barocas DA, Morgan TM, Resnick MJ, You C, et al. Incidence and predictors of understaging in patients with clinical T1 urothelial carcinoma undergoing radical cystectomy. *BJU Int.* 2014;113:894-9.
7. Kulkarni GS, Hakenberg OW, Gschwend JE, Thalmann G, Kassouf W, Kamat A, et al. An updated critical analysis of the treatment strategy for newly diagnosed high-grade T1 (previously T1G3) bladder cancer. *Eur Urol.* 2010;57:60-70.
8. Svatek RS, Hollenbeck BK, Holmang S, Lee R, Kim SP, Stenzl A, et al. The economics of bladder cancer: costs and considerations of caring for this disease. *Eur Urol.* 2014;66:253-62.
9. Babjuk M, Bohle A, Burger M, Capoun O, Cohen D, Comperat EM, et al. EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: Update 2016. *Eur Urol.* 2017;71:447-61.
10. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell.* 2017;171:540-56 e25.

11. Baylin SB, Jones PA. Epigenetic determinants of cancer. *Cold Spring Harb Perspect Biol.* 2016;8(9).
12. Jeronimo C, Henrique R. Epigenetic biomarkers in urological tumors: A systematic review. *Cancer Lett.* 2014;342:264-74.
13. Cooper DN. Eukaryotic DNA methylation. *Hum Genet.* 1983;64:315-33.
14. Weisenberger DJ. Characterizing DNA methylation alterations from The Cancer Genome Atlas. *J Clin Invest.* 2014;124:17-23.
15. van der Heijden AG, Mengual L, Ingelmo-Torres M, Lozano JJ, van Rijt-van de Westerlo CCM, Baixauli M, et al. Urine cell-based DNA methylation classifier for monitoring bladder cancer. *Clin Epigenetics.* 2018;10:71.
16. Su SF, de Castro Abreu AL, Chihara Y, Tsai Y, Andreu-Vieyra C, Daneshmand S, et al. A panel of three markers hyper- and hypomethylated in urine sediments accurately predicts bladder cancer recurrence. *Clin Cancer Res.* 2014;20:1978-89.
17. Feber A, Dhimi P, Dong L, de Winter P, Tan WS, Martinez-Fernandez M, et al. UroMark-a urinary biomarker assay for the detection of bladder cancer. *Clin Epigenetics.* 2017;9:8.
18. Ellinger J, Muller SC, Dietrich D. Epigenetic biomarkers in the blood of patients with urological malignancies. *Expert Rev Mol Diagn.* 2015;15:505-16.
19. Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans Neural Netw Learn Syst.* 2018;29:1774-85.
20. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010;26:1572-3.
21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284-7.
22. Zhu K, Liu Q, Zhou Y, Tao C, Zhao Z, Sun J, et al. Oncogenes and tumor suppressor genes: comparative genomics and network perspectives. *BMC Genomics.* 2015;16 Suppl 7:S8.
23. Chung CJ, Lee HL, Chang CH, Chang H, Liu CS, Jung WT, et al. Measurement of urinary arsenic profiles and DNA hypomethylation in a case-control study of urothelial carcinoma. *Arch Toxicol.* 2019;93:2155-64.
24. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer.* 2011;11:726-34.
25. Wolff EM, Chihara Y, Pan F, Weisenberger DJ, Siegmund KD, Sugano K, et al. Unique DNA methylation patterns distinguish noninvasive and invasive urothelial cancers and establish an epigenetic field defect in premalignant tissue. *Cancer Res.* 2010;70:8169-78.
26. Majewski T, Yao H, Bondaruk J, Chung W, Lee S, Lee JG, et al. Whole-organ genomic characterization of mucosal field effects initiating bladder carcinogenesis. *Cell Rep.* 2019;26:2241-56 e4.
27. Guan B, Xing Y, Xiong G, Cao Z, Fang D, Li Y, et al. Predictive value of gene methylation for second recurrence following surgical treatment of first bladder recurrence of a primary upper-tract urothelial carcinoma. *Oncol Lett.* 2018;15:9397-405.

28. Olkhov-Mitsel E, Savio AJ, Kron KJ, Pethe VV, Hermanns T, Fleshner NE, et al. Epigenome-wide DNA methylation profiling identifies differential methylation biomarkers in high-grade bladder cancer. *Transl Oncol.* 2017;10:168-77.
29. Yang Z, Liu A, Xiong Q, Xue Y, Liu F, Zeng S, et al. Prognostic value of differentially methylated gene profiles in bladder cancer. *J Cell Physiol.* 2019;234:18763-72.
30. Kitchen MO, Bryan RT, Haworth KE, Emes RD, Luscombe C, Gommersall L, et al. Methylation of HOXA9 and ISL1 predicts patient outcome in high-grade non-invasive bladder cancer. *PLoS One.* 2015;10:e0137003.
31. Xylinas E, Hassler MR, Zhuang D, Krzywinski M, Erdem Z, Robinson BD, et al. An epigenomic approach to improving response to neoadjuvant cisplatin chemotherapy in bladder cancer. *Biomolecules.* 2016;6.
32. Stubendorff B, Wilhelm K, Posselt K, Catto J, Hartmann A, Bertz S, et al. A three-gene methylation marker panel for the nodal metastatic risk assessment of muscle-invasive bladder cancer. *J Cancer Res Clin Oncol.* 2019;145:811-20.
33. Yoo CB, Jones PA. Epigenetic therapy of cancer: past, present and future. *Nat Rev Drug Discov.* 2006;5:37-50.
34. Kubicek S, O'Sullivan RJ, August EM, Hickey ER, Zhang Q, Teodoro ML, et al. Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Mol Cell.* 2007;25:473-81.
35. Segovia C, San Jose-Eneriz E, Munera-Maravilla E, Martinez-Fernandez M, Garate L, Miranda E, et al. Inhibition of a G9a/DNMT network triggers immune-mediated bladder cancer regression. *Nat Med.* 2019;25:1073-81.
36. Zhang H, Qi F, Cao Y, Zu X, Chen M, Li Z, et al. 5-Aza-2'-deoxycytidine enhances maspin expression and inhibits proliferation, migration, and invasion of the bladder cancer T24 cell line. *Cancer Biother Radiopharm.* 2013;28:343-50.
37. Carmack AJ, Soloway MS. The diagnosis and staging of bladder cancer: from RBCs to TURs. *Urology.* 2006;67:3-8; discussion 8-10.
38. Heitzer E, Perakis S, Geigl JB, Speicher MR. The potential of liquid biopsies for the early detection of cancer. *NPJ Precis Oncol.* 2017;1:36.
39. Roperch JP, Grandchamp B, Desgrandchamps F, Mongiat-Artus P, Ravery V, Ouzaid I, et al. Promoter hypermethylation of HS3ST2, SEPTIN9 and SLIT2 combined with FGFR3 mutations as a sensitive/specific urinary assay for diagnosis and surveillance in patients with low or high-risk non-muscle-invasive bladder cancer. *BMC Cancer.* 2016;16:704.
40. Chen F, Huang T, Ren Y, Wei J, Lou Z, Wang X, et al. Clinical significance of CDH13 promoter methylation as a biomarker for bladder cancer: a meta-analysis. *BMC Urol.* 2016;16:52.
41. Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. Cancer epigenetics: Moving forward. *PLoS Genet.* 2018;14:e1007362.
42. Park JW, Han JW. Targeting epigenetics for cancer therapy. *Arch Pharm Res.* 2019;42:159-70.

43. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. *Nat Rev Genet.* 2016;17:630-41.
44. Kiemeny LA, Grotenhuis AJ, Vermeulen SH, Wu X. Genome-wide association studies in bladder cancer: first results and potential relevance. *Curr Opin Urol.* 2009;19:540-6.
45. Wu X, Ye Y, Kiemeny LA, Sulem P, Rafnar T, Matullo G, et al. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat Genet.* 2009;41:991-5.
46. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev.* 2009;19:212-9.
47. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144:646-74.
48. Whongsiri P, Pimratana C, Wijitsettakul U, Sanpavat A, Jindatip D, Hoffmann MJ, et al. Oxidative stress and LINE-1 reactivation in bladder cancer are epigenetically linked through active chromatin formation. *Free Radic Biol Med.* 2019;134:419-28.
49. Chung CJ, Chang CH, Chuu CP, Yang CR, Chang YH, Huang CP, et al. Reduced 5-methylcytosine level as a potential progression predictor in patients with T1 or non-invasive urothelial carcinoma. *Int J Mol Sci.* 2014;16:677-90.
50. Kim S, Kim Y, Kong J, Kim E, Choi JH, Yuk HD, et al. Epigenetic regulation of mammalian Hedgehog signaling to the stroma determines the molecular subtype of bladder cancer. *Elife.* 2019;8.
51. Porten SP. Epigenetic alterations in bladder cancer. *Curr Urol Rep.* 2018;19:102.

Figures

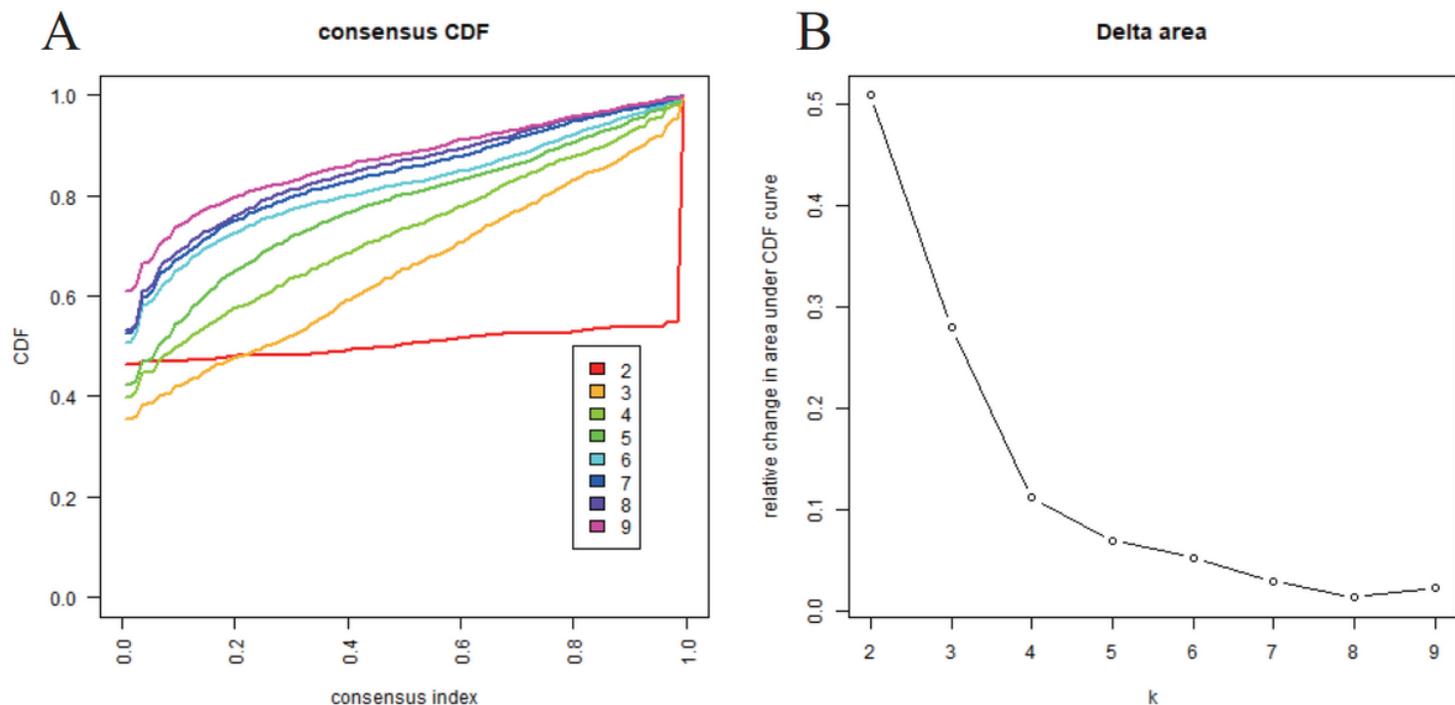


Figure 1

Consensus clustering of distinct bladder cancer DNA methylation prognostic subgroups. a. Cumulative distribution function (CDF) curve. b. CDF delta area curve. The delta area curve of consensus clustering shows the relative change in the area under the CDF curve for each category number k compared to that of k-1. The horizontal axis represents the category number k, and the vertical axis represents the relative change in the area under the CDF curve.

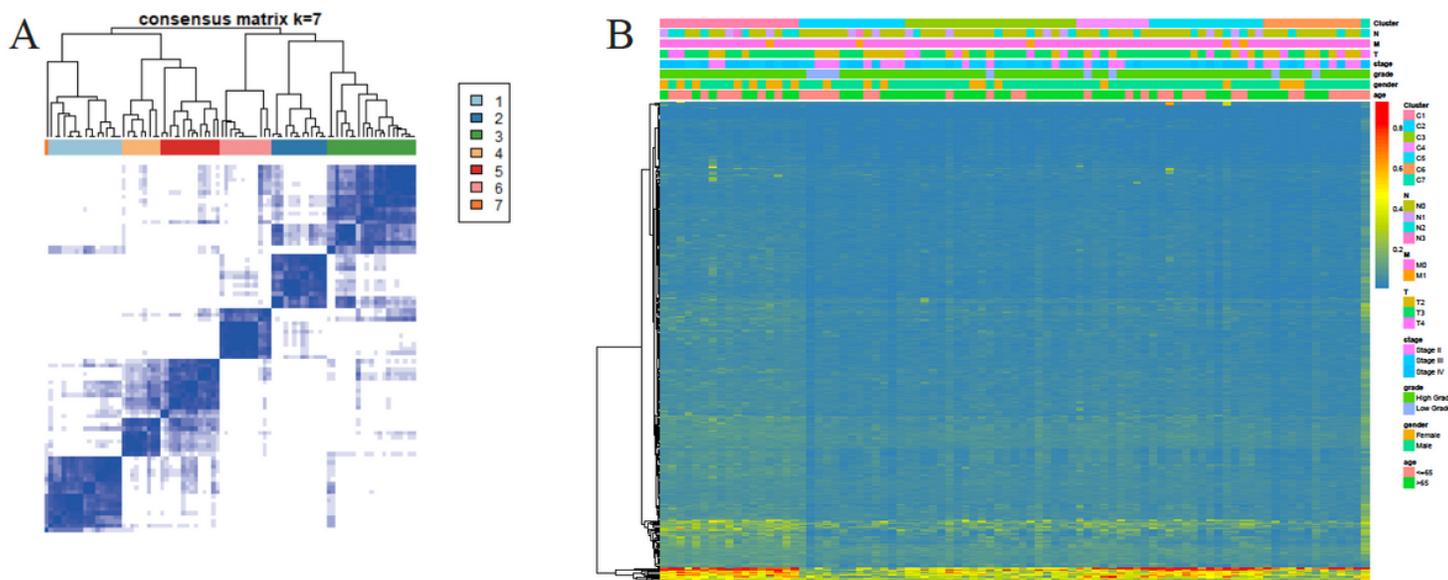


Figure 2

Cluster analysis of seven molecular subtypes by DNA methylation classification with the corresponding heat maps. a. Color-coded heat map corresponding to the consensus matrix of seven molecular subtypes, obtained by applying consensus clustering. b. The function of the heatmap annotated based on DNA methylation classification, TNM staging, clinicopathological staging, and histological type.

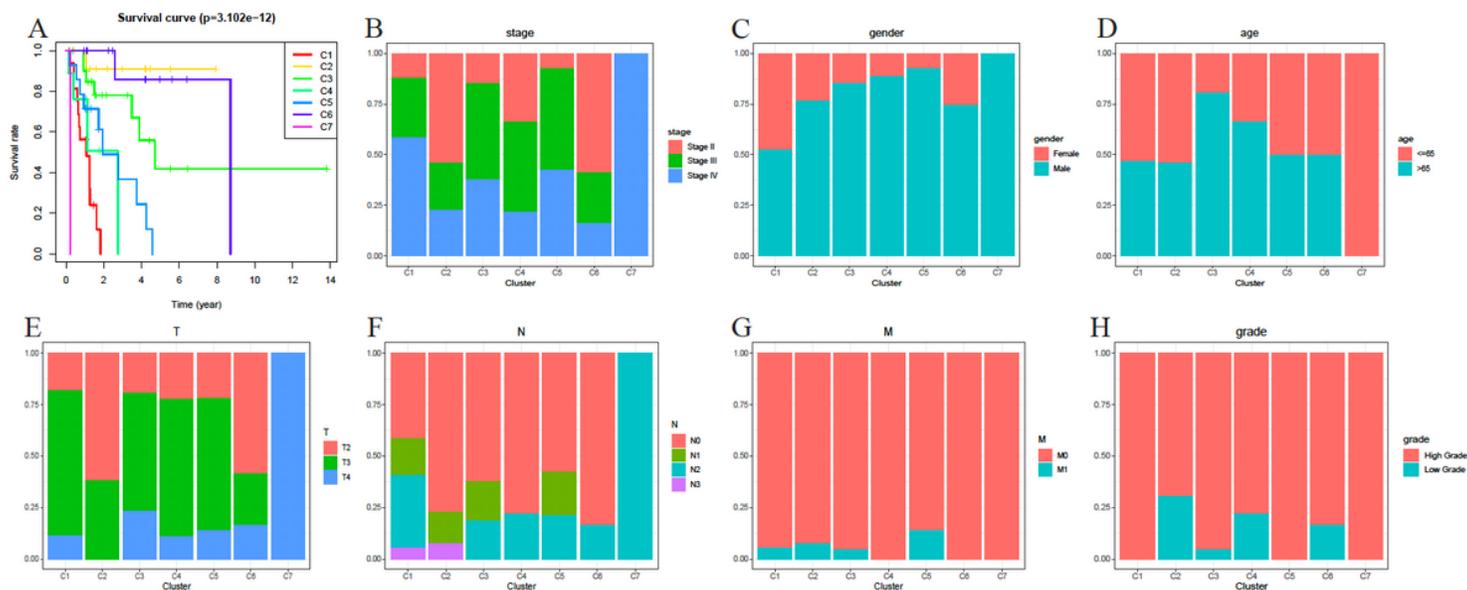


Figure 3

Comparison of prognosis, TNM stage, sex, age, and grade among the DNA methylation clusters. a. Survival curves for each DNA methylation subtype. b. Proportion of each clinical stage among the seven clusters. c. Proportion of both sexes among the seven clusters. d. Proportions of different ages among the seven clusters. e. Proportions of different T stage degrees among the seven clusters. f. Proportions of different N stage degrees among the seven clusters. g. Proportions of different M stage degrees among the seven clusters. h. Proportions of different pathological grades among the seven clusters.

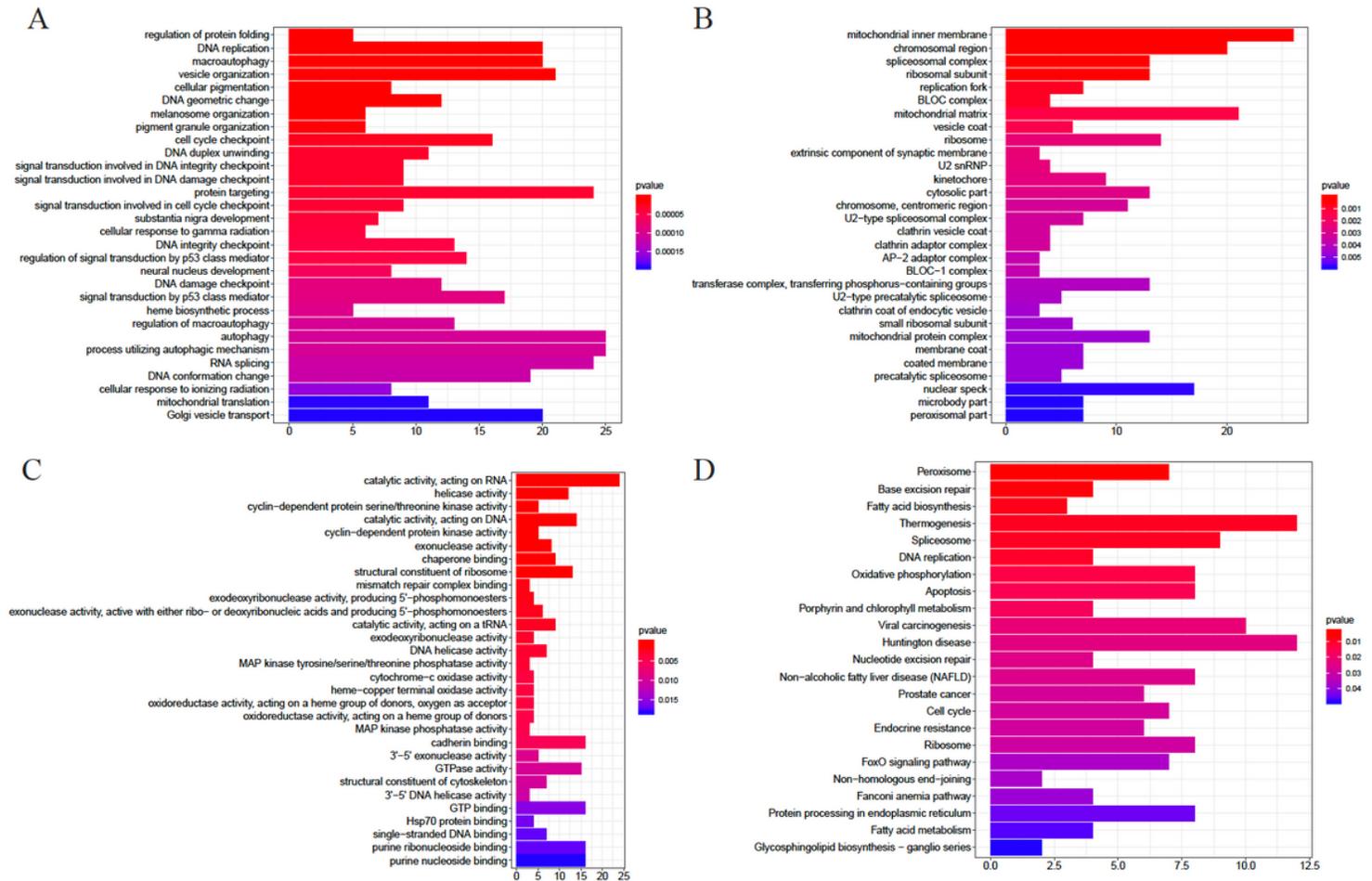


Figure 4

Enriched GO terms and KEGG pathways. a. GO terms for biological process. b. GO terms for cellular component. c. GO terms for molecular function d. KEGG pathways.

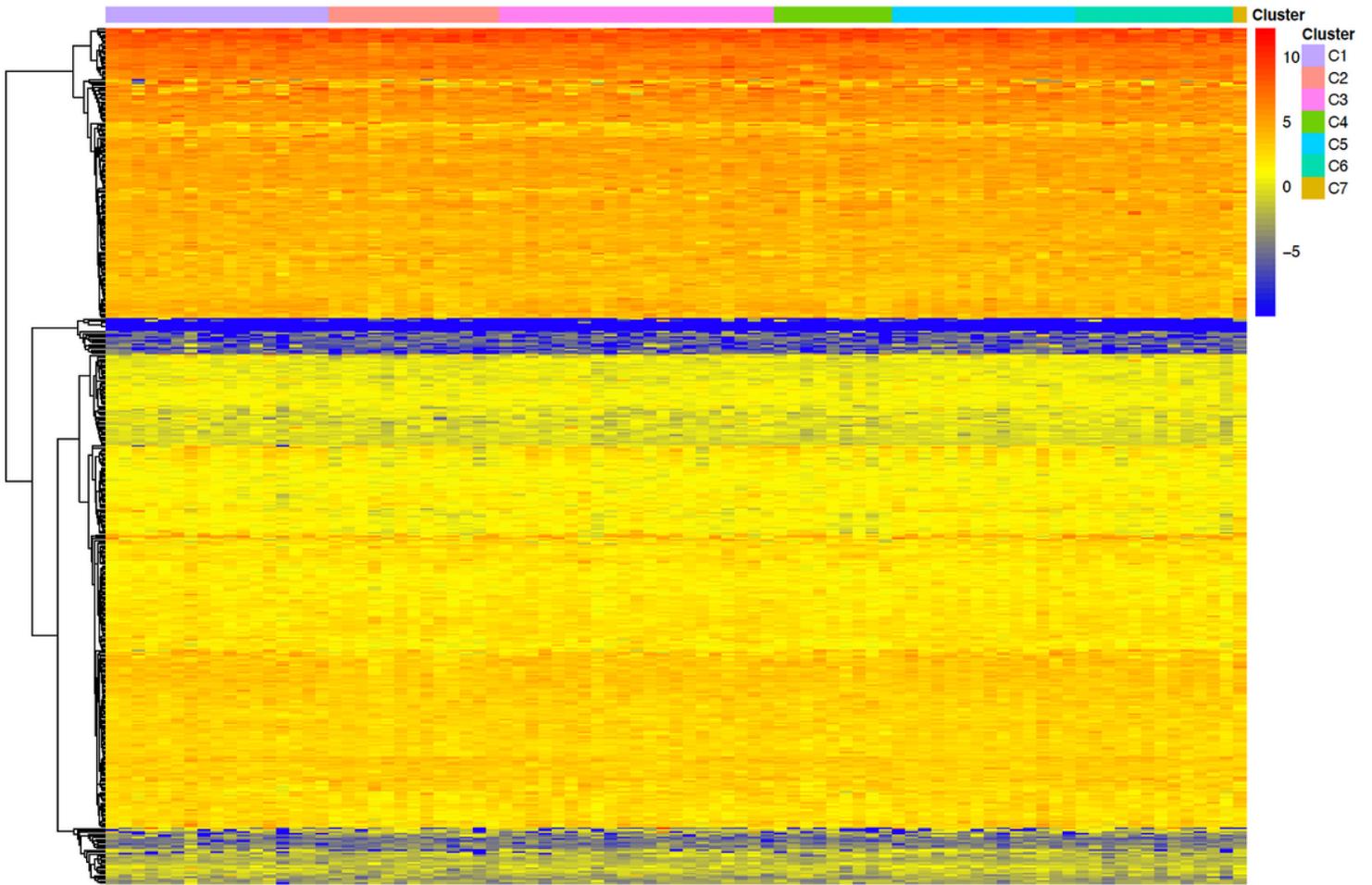


Figure 5

Heatmap of annotated genes associated with the 456 CpGs.

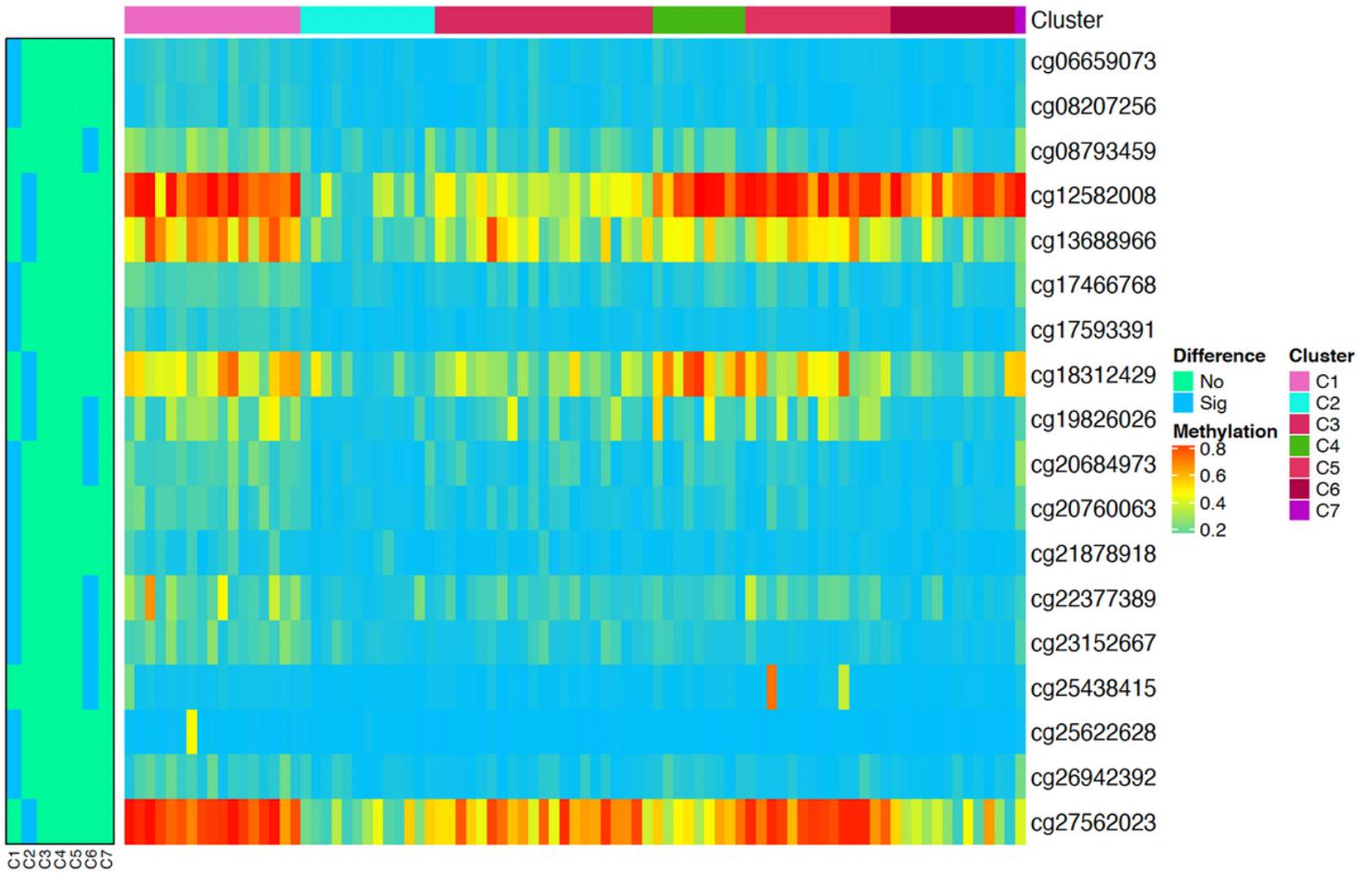


Figure 6

Specific methylation of CpGs for each DNA methylation cluster. Specific CpGs are shown for each DNA methylation prognostic subtype. Red and blue represent hyper- and hypomethylated CpGs, respectively.

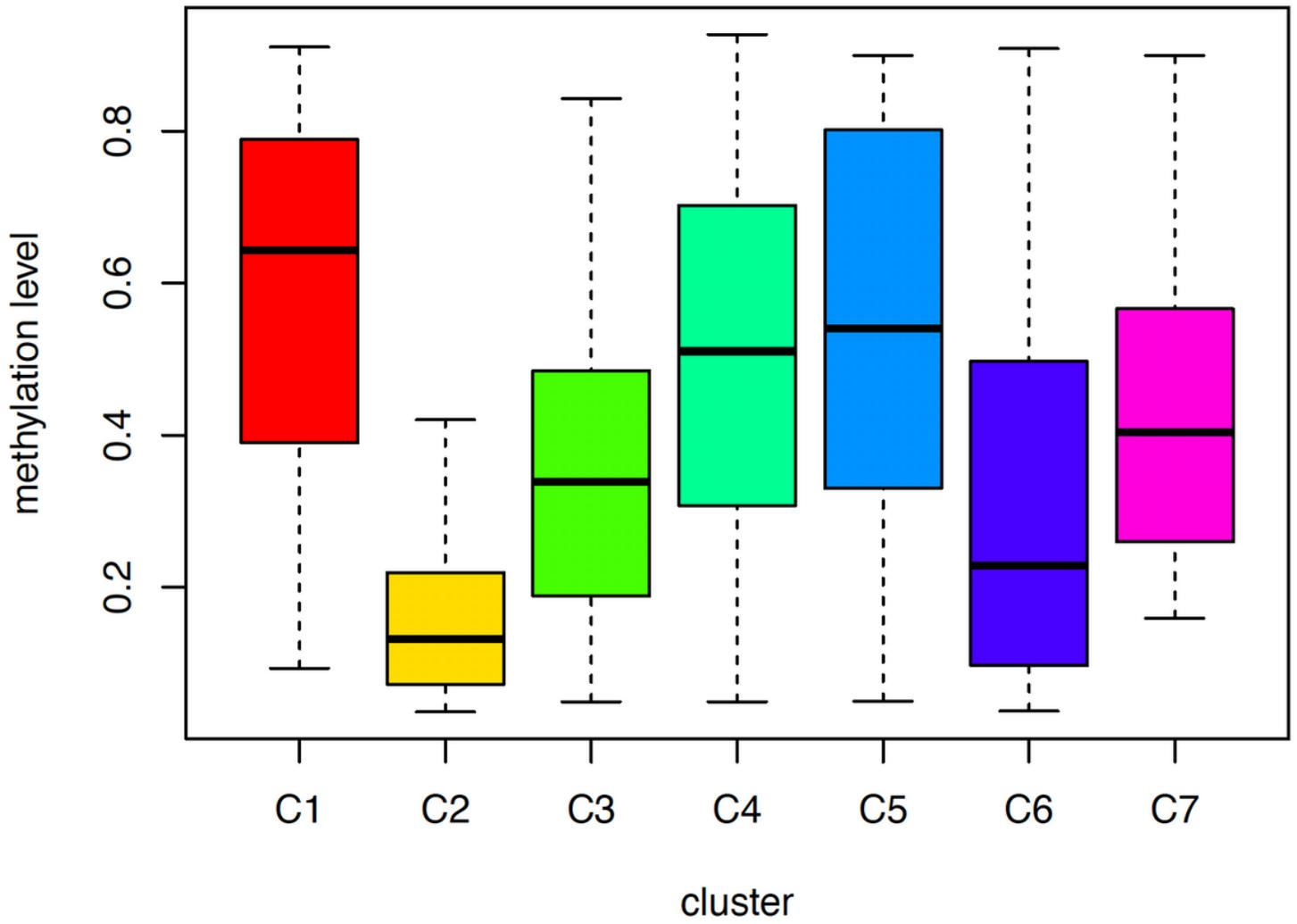


Figure 7

Box plot of CpG methylation levels in the seven clusters.

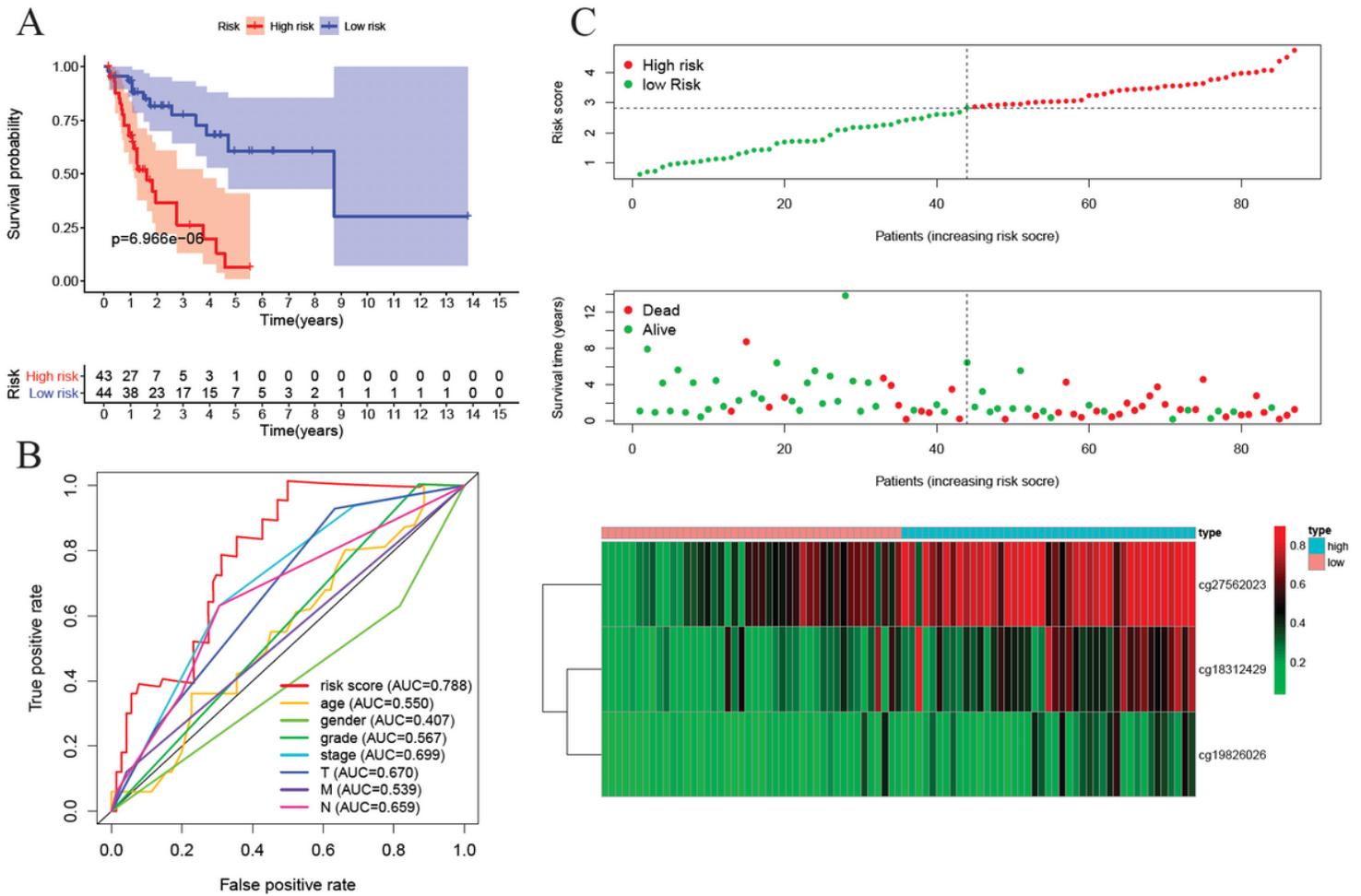


Figure 8

Construction of the prognostic prediction model for the training group. a. Kaplan–Meier survival analysis of high- and low-risk groups in the training group. b. Time-dependent receiver operation characteristic (ROC) of the indicated predictive factors in the training group. c. Rank of risk score and distribution of groups, survival status of patients in different groups, and expression heatmap of the three CpGs sites included.

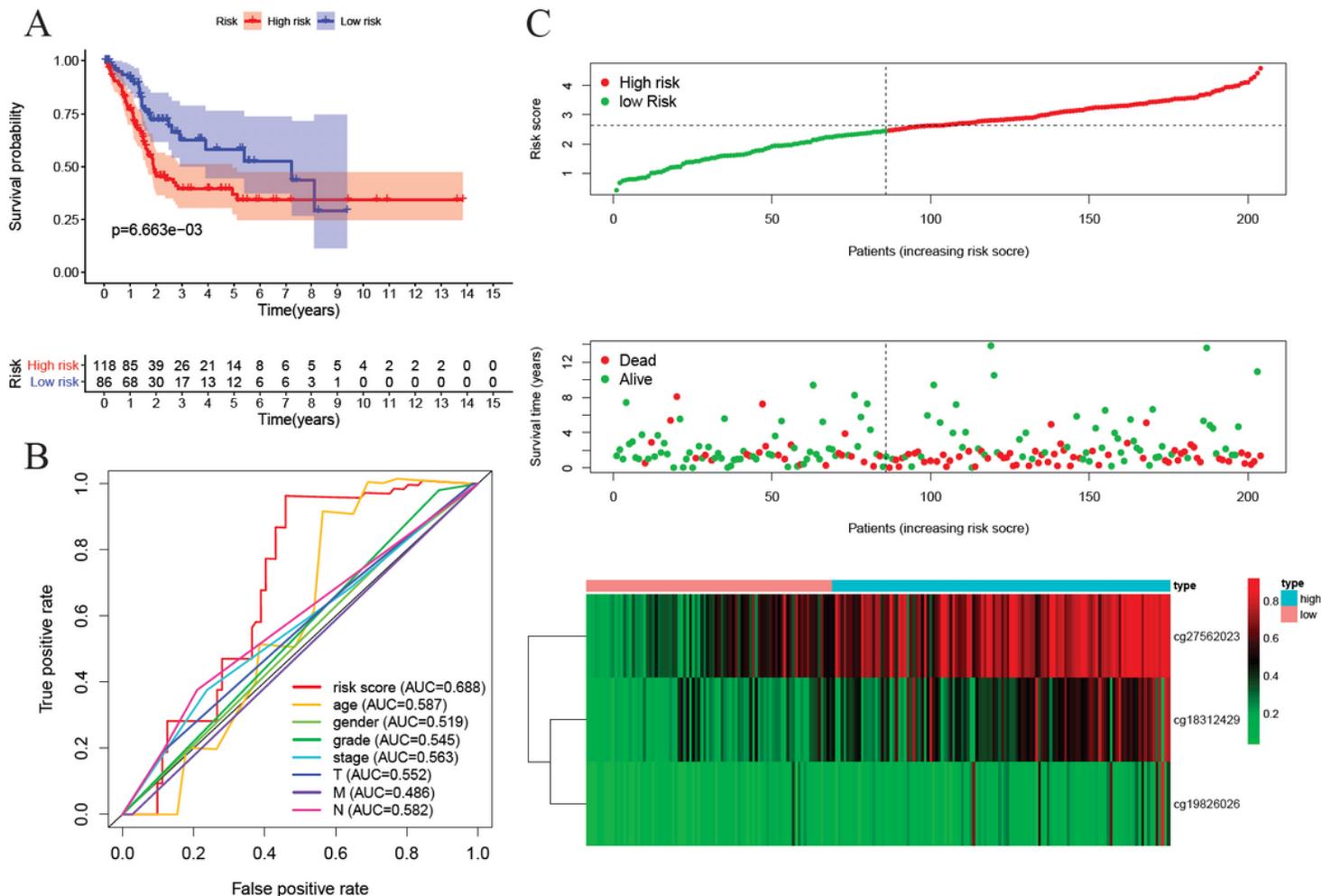


Figure 9

Construction of the prognostic prediction model for the testing group. a. Kaplan–Meier survival analysis of high- and low-risk groups in the testing group. b. Time-dependent receiver operation characteristic (ROC) of the indicated predictive factors in the testing group. c. Rank of risk score and distribution of groups, survival status of patients in different groups, and expression heatmap of the three CpGs sites included.

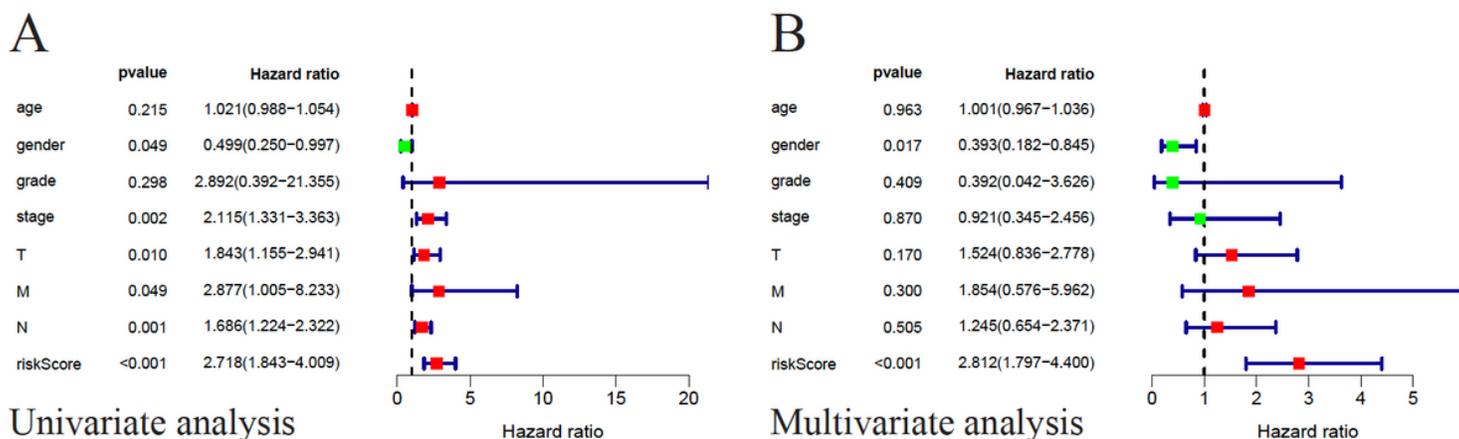


Figure 10

Univariate and multivariate COX regression analyses of bladder cancer cohort. Results showed that sex and the prognostic model were independent factors for predicting the overall survival rate of bladder cancer patients.

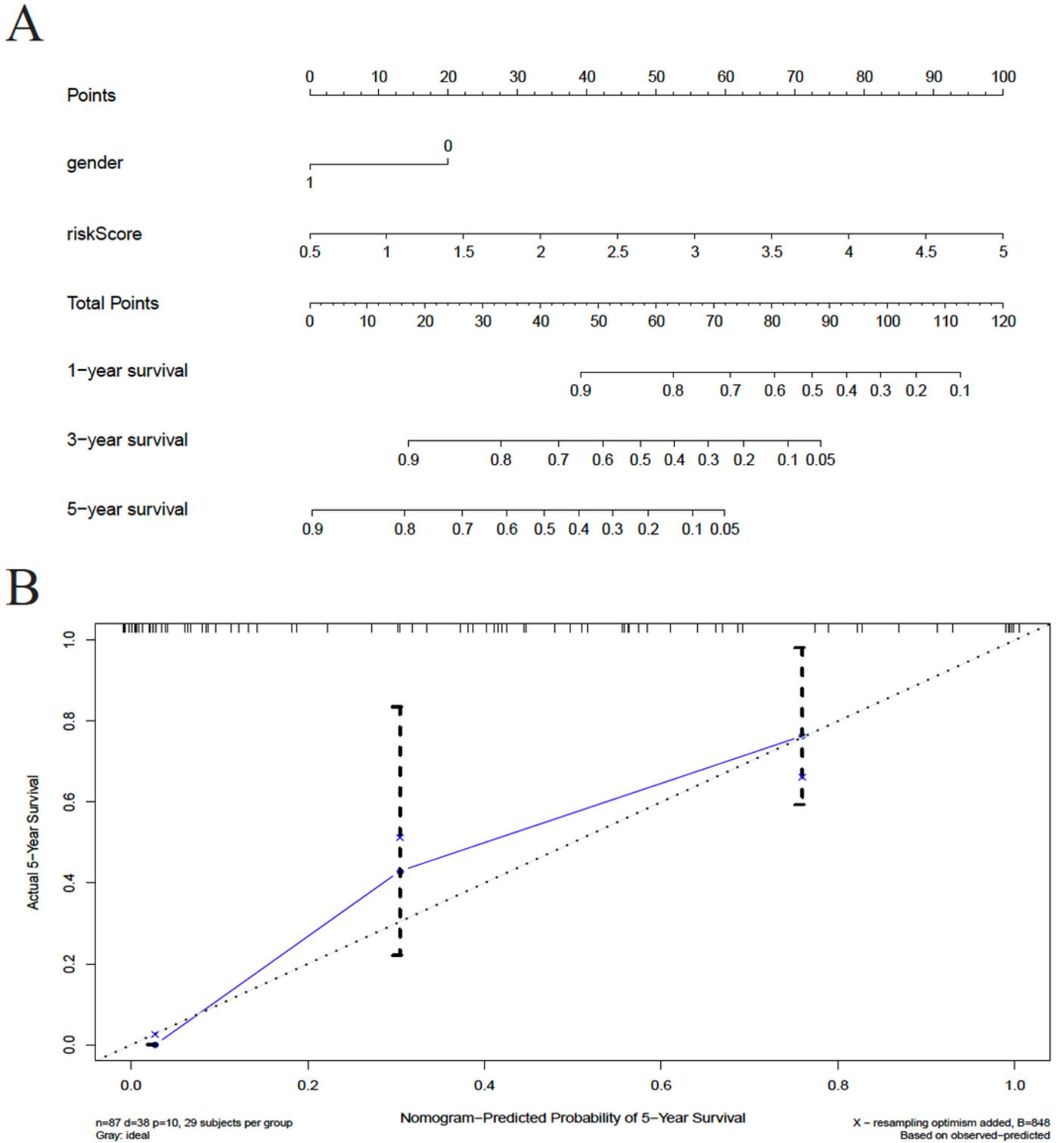


Figure 11

Nomogram and calibration plots a. Nomogram to predict the probability of overall survival (OS) in patients with bladder cancer. b. Calibration plots for the nomogram at 5 years.

Supplementary Files

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

- [Supp.Table1.docx](#)