

Prediction of Prognosis in Patients with Prostate Cancer by DNA Methylation Classification

Lei Chen

Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Deshen Pan

Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Minglei Sha

Department of Geriatric, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Deng Li

Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Chaoliang Xu

Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Feiteng Chen

Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Jiayu Liu

Hangzhou Normal University, School of Medicine, Hangzhou, China

Danchuan Zheng

Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Jun Lu

Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Shujie Xia

Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Yi Shao (✉ drshaoyi@163.com)

Shanghai First People's Hospital <https://orcid.org/0000-0002-5842-3167>

Research

Keywords: Prostate cancer, DNA methylation, Prognosis, Prediction

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Posted Date: July 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-38332/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Prostate cancer is the second most frequently diagnosed cancer and the fifth leading cause of cancer-related death. It is estimated that the incidence of prostate cancer is on the rise worldwide. Epigenetic changes in tumors play an important role in the occurrence and development of prostate cancer. DNA methylation is one of the mechanisms of tumor epigenetic regulation and may be a new biomarker that has great potential in early tumor screening, treatment guidance and prognosis prediction. The purpose of this study was to explore a classification method from the perspective of DNA methylation.

Methods: The least absolute shrinkage and selection operator (LASSO) method was used to analyze DNA methylation and RNA-seq data from the Cancer Genome Atlas (TCGA). The methylation sites with small differences were eliminated, and the 21 methylation sites with the most significant differences were retained for analysis. Using their corresponding gene expression levels, a recurrence prediction model for prostate cancer patients was constructed to distinguish high-risk, medium-risk, and low-risk cases. Immune cell abundance analysis, gene enrichment analysis, Tumor burden mutation analysis and gene copy number variation analysis were then used to analyze the differences among these three subtypes and their underlying mechanisms.

Results: We observed the difference in disease-free survival (DFS) of the three methylated subtypes in the test set, which was verified in the validation set. We found three subtypes have different proportions of immune cells, especially in memory B cells, M2 macrophages, Treg cells. GSVA and GSEA analysis revealed that the relevant metastasis gene sets of prostate cancer were enriched in high-risk cases. In addition, the mutation frequencies of *TP53*, *TTN* and *KMT2D* were the highest, and gradually increased in the three genotypes according to Tumor burden mutation (TMB) analysis. Gene copy number variation (CNV) showed that *AR*, *LAPTM4B*, and *MTDH* were significantly amplified, while *ATP1B2* and *FAM92B* were significantly deleted. Finally, univariate and multivariate analysis showed that there were statistical differences among the three methylation subtypes, which can be used as an index to predict prostate cancer recurrence.

Conclusions: Our study suggests that classification based on DNA methylation is an independent factor for predicting tumor recurrence in patients with prostate cancer.

Background

Prostate cancer is the second most commonly diagnosed cancer (1,276,106 new cases in 2018, 13.5% of all cancer cases in men) and the fifth leading cause of cancer-related death (358,989 deaths in 2018, 3.8% of all deaths caused by cancer in men) in males around the world[1]. It is estimated that the global incidence of prostate cancer is on the rise, with 1,017,712 new cases (+ 79.7% overall change) projected through 2040[2, 3]. Increasing incidence seems to be related to extending life expectancy. In developing countries, it may also be due to increased access to medical care and increased case records and reports. Research has shown that genetic background and environmental factors play an important role in the development of prostate cancer[2].

In recent years, epigenetic modifications were found to play an important role in tumorigenesis. Epigenetic inheritance is defined as cellular information, other than the DNA sequence itself, that is heritable during cell division. Epigenetics mainly include DNA methylation, genomic imprinting and histone modification, the most common of which is DNA methylation[4]. There are two types of DNA methylation: hypermethylation, which inhibits gene expression, and hypomethylation, which activates gene expression. Functionally, tumor suppressor genes are frequently hypermethylated, and oncogenes are hypomethylated. To date, dozens of common abnormally methylated genes have been identified in prostate cancer. These genes encompass many cellular functions, including cell cycle control, apoptosis, hormone response, DNA repair and damage prevention, signal transduction, tumor invasion and tumor suppression[5]. Frequent promoter methylation of genes such as glutathione S-transferase π (GSTP1)[6] metastasis suppressor 1 (MTSS1)[7] dedicator of cytokinesis 2 (DOCK2)[8] adenomatous polyposis coli (APC) [9] and retinoic acid receptor beta2(RAR β 2) [10] has been found and is associated with the progression of prostate cancer. However, whether the methylation of these genes is related to important clinical factors, such as the degree of tumor malignancy and biochemical recurrence and prognosis, has not been determined from large datasets of patients with prostate cancer. Therefore, this study was designed to explore a classification method from the perspective of DNA methylation and evaluate its potential for clinical application to help clinicians evaluate efficacy and prognosis and guide treatment decisions.

Materials And Methods

1. Data acquisition and data analysis

The DNA methylation data of prostate cancer patients were downloaded from the TCGA database [11], and methylation 450K microarray data, RNA-seq data and related clinical follow-up data of prostate cancer patients were obtained. Only the data with clinical survival information were selected for the analysis of the correlation between methylation level and disease-free survival. Then, the CpG site with NA (not available) and the cross-reaction CpG site in the genome were deleted [12]. In addition, the k-nearest neighbor method in the impute R package for methylation profiles was utilized to impute the missing values, with the further removal of the unstable genomic methylation sites[13]. In other words, the CpG sites and single nucleotide sites from the sex chromosomes were discarded, and the difference between the cancer tissue sample group and the adjacent normal (AN) prostate tissue sample group was analyzed by using the CHAMP package of R software. Specifically, we obtained measurement data for each CpG site, including methylated signal intensity (denoted by M) and unmethylated signal intensity (denoted by N). Methylation status was expressed as the β value ($\beta = M/(M + U)$, ranging from 0 to 1). In total, 215166 differential DNA methylation sites were selected by setting $p < 0.01$ and $p.adjust < 0.01$. Finally, 5406 differential DNA methylation sites were obtained by further screening under the conditions of average β value of CpG sites in AN samples < 0.2 and SD (standard deviation) of CpG sites in tumors > 0.2 .

The whole exome sequencing (WES) data of 432 samples were downloaded from TCGA. WES data provide a spectrum of somatic mutations in patients with prostate cancer. Tumor mutational burden (TMB), a new biomarker of immunotherapy response, is defined as the number of nonsynonymous mutations per trillion

bases[14]. The difference in total mutational burden in different subtypes was obtained by calculating the number of nonsynonymous mutations in the whole genome and analyzing them by the Kruskal-Wallis method.

The level 4 copy number variation (CNV) profiles of prostate cancer patients were downloaded from TCGA and classified according to methylation subtypes. The amplification or deletion changes in the whole genome were identified by the GISTIC tool, the amplification or deletion gene segments were further screened by setting the Q value > 0.25, and gene annotations were carried out to observe the changes in CNV in methylation subtypes.

Complete clinical information of prostate cancer patients in the TCGA cohort, including gleason score, clinical M, pathologic N, pathologic T, biochemical recurrence, treatment outcome and age, was collected and used for follow-up analysis.

The GSE2612 dataset based on the GEO platform contains prostate cancer methylation data, RNA-seq data, clinical follow-up data and complete clinical information and was downloaded and used for subsequent analysis.

2. LASSO regularization, signature construction and methylation subtype classification

LASSO (least absolute shrinkage and selection operator) is a method of linear model estimation using L1 regularization,[15] which constrains the coefficient estimation to the direction of 0 so that some independent variables are eliminated and sparse feature space is generated. We used the glmnet R software package for LASSO regression analysis. Twenty-one genes are selected to construct the signature, and each coefficient is obtained through the constraint process. The risk score formula is as follows: risk score = $\sum_i \omega_i \chi_i$ where ω_i and χ_i are the coefficient of each gene and the expression value of each gene, respectively. The names of the genes and their corresponding coefficients are shown in Table 1. X-tile software was used to calculate the optimal cutoff value and to classify methylation subtypes. The overall survival rates (the percentage of people in a group who were alive after a length of time) and disease-free survival rates (the percentage of people in a group who were disease-free after a length of time) of these subtypes were compared using the Survival R software package based on the log-rank (Mantel-Cox) test.

3. Analysis of immune cell abundance

The tumor infiltration fractions of various immune cell subtypes were calculated by the CIBERSORT method. In this study, a deconvolution algorithm was utilized to distinguish 22 immune cell subtypes, including dendritic cells, NK cells, neutrophils, eosinophils, mast cells, monocytes, macrophages and several B cell and T cell types in adaptive immunity. ANOVA was performed to evaluate the differences of 22 human immunocyte phenotypes in tumor infiltration.

4. Functional enrichment analysis of the gene set

To evaluate the potential mechanism of methylation typing in prostate cancer, GSVA analysis and ANOVA were performed to identify the differences in pathways and gene sets in different subtypes of prostate cancer ($P < 0.01$). For cluster 3, which has a poor prognosis, we further carried out GSEA analysis to reveal the mechanism related to poor prognosis.

5. Other statistical analyses

The ability of factors to predict 1-year and 3-year recurrence rates was tested using ROC curves. The relationship between clinical features and methylation subtypes was analyzed by Fisher's precise test. Methylation typing and clinical features were analyzed by univariate and multivariate Cox analysis.

Results

1. Identification of differential methylation sites

To identify significant DNA methylation sites in prostate cancer, we downloaded the data from methylation 450K chips of 502 prostate cancer (PC) and 50 adjacent normal (AN) prostate tissue samples from TCGA. The ChAMP package in R software was utilized to analyze the differentially methylated CpG sites between PC and AN samples, and 215,166 methylation sites were screened out with the significance thresholds set as $p < 0.01$ and $p.adjust < 0.01$ (Table S1). The DNA hypermethylation sites in prostate cancer were further identified under the conditions of average β value of CpG sites in AN samples < 0.2 and SD of CpG sites in tumors > 0.2 . As a result, we obtained 5406 differential DNA methylation sites for further analysis.

2. LASSO Cox regression identifying a 21-gene epigenetic signature

The large number of methylation sites is not conducive to clinical application. To filter the CpG sites most relevant for clinical recurrence while maintaining effectiveness, the LASSO Cox model was applied to the gene expression profile corresponding to the 5406 CpG sites. The LASSO algorithm is a method of linear model estimation using L1 regularization. It constrains the coefficient estimate toward 0 so that some independent variables are eliminated, thereby achieving the purpose of simplifying the model and selecting characteristic factors. In this study, the glmnet R software package was used for LASSO regression analysis. The coefficient trajectory of each independent variable is shown in Fig. 1a. As the penalty term coefficient λ increases, the independent variable coefficients are gradually constrained. A 10-cross validation was used to analyze the confidence interval (Fig. 1b) at each lambda value. To minimize the error, we found the best penalty term coefficient, which contained 21 genes (Table 1).

Table 1
Names of 21 genes and their corresponding CpG sites and constraint coefficients

CpG sites ID	coefficient	Gene name
cg00582971	0.013972545	GRM6
cg01135464	0.061856371	TOB1-AS1
cg04828133	1.23E-08	-
cg04844977	-0.449669865	ROPN1B
cg04949489	-0.011021213	LDHB
cg04998420	0.143078025	S100A6
cg06223466	-0.327813667	RADIL
cg08572315	-0.908578589	TRABD2A
cg08814105	0.40040525	NKX2-5
cg09474331	0.096565122	TTYH1
cg12054869	-9.98E-09	N4BP2L1
cg12574408	-1.53E-09	C7orf31
cg13442969	-0.213755758	DYRK2
cg14249876	0.03810652	AEBP1
cg14499577	-0.085621717	-
cg14516385	0.232145878	BX842568.4
cg14699728	0.393343407	NPAS4
cg15415945	0.358192218	C6orf47
cg17030173	1.65E-09	MARCH11
cg18310412	1.36E-08	TRIM71
cg23305567	0.195833948	AC008753.6

3. Risk classification of recurrence based on epigenetic signature

The coefficients and gene expression values of 21 genes were used to calculate the risk score of each sample (formula in the method section) to construct a recurrence-related model. To obtain DNA methylation subtypes, X-tile software was used to calculate the best cutoff value and classify patients into three groups:

the subsequent figures, respectively denoted by

cluster 1, cluster 2 and cluster 3) (Fig. 2a). The methylation difference heat map of the three groups is shown in Fig. 2b. To determine the effect of this classification method, the survival R software package was used to draw the disease-free survival (DFS) curve, which showed the clinical recurrence probability of different groups. We found that the high-risk group was the most likely to relapse, and the low-risk group was the least likely to relapse ($p < 0.0001$) (Fig. 2c). We used the same method to draw the overall survival curves, and we found that the survival time of the high-risk group was significantly shorter than that of the low-risk group ($p = 0.036$) (Fig. 2d).

In addition, we found that there were significant differences in the expression of androgen receptors (AR) among different subtypes ($p = 0.013 < 0.05$) (Fig. 2e). Pairwise comparative analysis showed that there were significant differences in the expression of AR between cluster 1 and cluster 3 ($p = 0.04 < 0.05$), cluster 2 and cluster 3 ($p = 0.036 < 0.05$). It means methylation classification was highly associated with the expression of AR, and the highest expression of AR was found in cluster 3 (high-risk group) among the three groups.

4. Differences in immune cell abundance among methylation subtypes

To observe the immunological characteristics of methylation subtypes, the CIBERSORT method was used to analyze the degree of infiltration of immune cells in tumor tissue samples. We found that samples with the three different methylation subtypes had different proportions of immune cells, especially memory B cells, naïve B cells, M2 macrophages, Treg cells, γ/δ T cells, plasma cells and mast cells (Fig. 3). No significant difference was found in dendritic cells, NK cells, neutrophils, eosinophils, monocytes or other types of T and B cells (Fig S1). For the cluster 3 group which has a poor prognosis, the proportion of memory B cells, M2 macrophages and Treg cells was significantly higher than that of cluster 1 and cluster 2. This is consistent with previous studies[16–18]. M2 macrophages and Treg cells are immunosuppressive cells that inhibit immune activation in tumors and promote the development of tumors. This indicates that the cluster 3 subtype has a unique immunophenotype, which is characterized by more immunosuppressive cell infiltration and lower immune activation. Different immunophenotypes of methylation subtypes may be the potential mechanism of different prognoses.

5. Difference analysis of pathways

To determine the biological processes related to these three types, we analyzed the TCGA-PRAD data by GSVA and generated a heat map (Fig. 4a). The results showed that among the three subtypes, some gene sets gradually increased, such as "REACTOME_SUMOYLATION_OF_DNA_METHYLATION_PROTEINS", which affected the level of DNA methylation, while other gene sets gradually decreased, such as "CHANDRAN_METASTASIS_TOP50_DN", which was a group of genes downregulated in metastatic prostate cancer. This gene set was highly expressed in cluster 1 and decreased gradually in cluster 2 and cluster 3. In line with the GSVA results, GSEA analysis showed that the gene set of "RAMASWAMY_METASTASIS_UP" was enriched while the gene sets of "CHANDRAN_METASTASIS_DN" and "ELLWOOD_MYC_TARGETS_DN" were down-regulated in patients of cluster 3 group (Fig. 4b-d). These results indicated the association between DNA methylation subtypes and tumor metastasis. In general, the aforementioned biological

methylation in the progression of prostate cancer.

6. Difference analysis of tumor mutational burden

Tumor burden mutation (TMB) is defined as the total number of nonsynonymous mutations per megabase. To investigate the differences among the three types of prostate cancer. Data from 432 samples, those with both WES data and the DNA methylation data, were analyzed, and the results showed that TMB was significantly different among the three types (Fig. 5a). The TMB increased gradually with risk (cluster 1 < cluster 2 < cluster 3), suggesting that the risk score was positively correlated with TMB. We further analyzed the mutation burden of different genes in different subtypes, and we found that the mutation frequencies of tumor protein p53 (TP53), titin (TTN) and lysine methyltransferase 2D (KMT2D) were the highest and gradually increased in the three genotypes (Fig. 5b).

7. CNV analysis

We next investigated gene copy number variation (CNV) in prostate cancer. The ratio of tumor tissue to normal tissue in each gene segment was obtained by analyzing the CNV probe in the Affy SNP6.0 chip. The GISTIC tool was utilized to further calculate and screen out regions that were significantly amplified and significantly deleted (Fig S2). Gene segments with Q values > 0.25 were selected for gene annotation to observe the CNV changes of prostate cancer in different subtypes (Fig. 5c) (see Tables S2 and S3 for genes contained in each chromosome segment). The results showed that the CNV amplification or deletion of genes was positively correlated with the risk score. For cluster 3 patients with poor prognosis, AR, lysosomal protein transmembrane 4 beta (LAPTM4B), and metadherin (MTDH) were significantly amplified, while ATPase Na⁺/K⁺ transporting subunit beta 2 (ATP1B2) and family with sequence similarity 92, member B (FAM92B) were significantly deleted.

8. Evaluation and validation of models

To evaluate the methylation model we devised, the 1-year and 3-year boundaries were set to observe the clinical recurrence of patients, and the ROC curve was drawn to evaluate its effectiveness in predicting recurrence (Fig. 6a). The accuracy of the model for predicting 1 year recurrence was slightly higher than that of for predicting 3-year recurrence. In general, the methylation model has good sensitivity and specificity for predicting whether patients will relapse within 1 or 3 years.

We used Fisher's exact test to analyze the association between clinical characteristics and methylation subtypes (Table 2). Clinical characteristics in 3 subtypes, including "Gleason score", "Clinical M", "Pathologic N", "Pathologic T", and "Biochemical recurrence", were found to be significantly different, while "Treatment outcome" was not. In addition, we performed univariate and multivariate relapse analyses of methylation subtypes and clinical characteristics (Table 3). Univariate analysis showed that "DNA methylation subtypes", "Gleason score", "Biochemical recurrence", "Treatment outcome", "Pathologic N", and "Pathologic T" all had significant differences, while multivariate analysis showed that "DNA methylation subtypes" had significant differences among relapsed and nonrelapsed patients. These data indicate that methylation typing can be used as an index to predict the recurrence of prostate cancer.

Table 2
Statistical differences between clinical characteristics and methylated subtypes

Clinical characteristics		cluster3	cluster2	cluster1	P value
Gleason score	< 8	2(4.26%)	59(55.66%)	206(71.78%)	< 2.2E-16
	>=8	45(95.74%)	47(44.34%)	81(28.22%)	
Clinical M	M0	38(84.44%)	100(99.01%)	256(99.22%)	1.27E-05
	M1	7(15.55%)	1(0.99%)	2(0.78%)	
Pathologic N	N0	24(53.33%)	75(80.65%)	217(90.04%)	0.04669
	N1	21(46.67%)	18(19.35%)	24(9.96%)	
Pathologic T	T2	3(6.52%)	31(29.81%)	144(50.88%)	5.10E-08
	T3 + T4	43(93.48%)	73(70.19%)	139(49.12%)	
Biochemical recurrence	NO	26(61.9%)	76(89.41%)	227(91.16%)	4.71E-06
	YES	16(38.1%)	9(10.59%)	22(8.84%)	
Treatment outcome	CR + PR + SD	29(90.63%)	46(93.88%)	118(95.16%)	0.4054
	PD	3(9.38%)	3(6.12%)	6(4.84%)	
Age	< 61	14(29.79%)	41(38.68%)	143(49.83%)	0.02984
	>=61	33(70.21%)	65(61.32%)	144(50.17%)	

Abbreviations: CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease.

Table 3

Univariate and multivariate relapse analyses of methylation typing and clinical characteristics in training set

Clinical characteristics	Univariable analysis				Multivariable analysis			
	HR	LOWER	HIGHER	P-VALUE	HR	LOWER	HIGHER	P-VALUE
DNA methylation subtypes	11.2	6.93	18.1	< 2.2E-16	6.77584	3.22204	14.249	4.54E-07
Gleason score	5.79	3.38	9.92	< 2.2E-16	1.88647	0.63726	5.585	0.2517
Biochemical recurrence	4.88	3.01	7.93	< 2.2E-16	1.43414	0.69325	2.967	0.331
Treatment outcome	2.52	1.2	5.29	0.014	1.64217	0.66481	4.056	0.2823
Age	1.45	0.93	2.27	0.104	1.09601	0.5323	2.257	0.8035
Clinical M	2.13	0.91	5	0.083	0.36246	0.09295	1.413	0.1438
Pathologic N	3.18	1.98	5.13	< 2.2E-16	1.84842	0.86926	3.931	0.1105
Pathologic T	9.11	3.93	21.13	< 2.2E-16	6.30546	0.76175	52.194	0.0877

Abbreviations: HR, hazard ratio.

Finally, the GSE26126 dataset in the GEO database was used to validate the model we built. We used 21 identified markers to calculate risk scores and perform recurrence analysis. The results showed that DNA methylation subtypes have significant differences in the probability of clinical recurrence (Fig. 6b). We found that the high-risk group (cluster 3) was the most likely to relapse, and the low-risk group (cluster 1) was the least likely to relapse ($p < 0.0001$), which is consistent with our previous results. Similarly, we performed univariate and multivariate relapse analyses using methylation subtypes and clinical characteristics (Table 4). The results showed that there were significant differences in DNA methylation subtypes in both the univariate analysis and the multivariate analysis. It turned out methylation typing is a universal index for predicting the recurrence of prostate cancer.

Table 4

Univariate and multivariate relapse analyses of methylation typing and clinical characteristics in validation set

Clinical characteristics	Univariable analysis				Multivariable analysis			
	HR	LOWER	HIGHER	P-VALUE	HR	LOWER	HIGHER	P-VALUE
DNA methylation subtypes	10.08	3.13	32.47	< 2.2E-16	11.15715	2.9042	42.863	0.000444
Age	0.68	0.25	1.83	0.441	0.97757	0.2981	3.206	0.970133
Pre-treatment PSA	3.47	1.25	9.64	0.017	3.46513	1.0952	10.963	0.034451
Gleason score	1.61	0.84	3.08	0.152	0.90812	0.4623	1.784	0.779644

Abbreviations: HR, hazard ratio.

Discussion

Epigenetics is a heritable phenotypic change that occurs without changes in the genomic DNA sequence. In normal biological processes, epigenetics alters the expression of genes, allowing cells to adapt to the surrounding environment and exert their unique functions. Epigenetic dysregulation is also involved in the processes of tumorigenesis, proliferation and metastasis, epigenetic changes in tumor cells. Abnormal gene expression and biological processes are conducive to tumor survival and facilitate characteristic alterations to tumors. There is great potential for the application of epigenetics in early tumor screening, treatment guidance and prognosis prediction. The main epigenetic modifications include DNA methylation, histone modification and genomic imprinting. As the most common epigenetic mechanism, DNA methylation is commonly altered in tumors. Hypomethylation of DNA leads to gene activation. Many CpG islands are usually methylated in somatic tissue. These methylated islands may become undermethylated in cancer, which can activate adjacent genes. Examples of genes affected by hypomethylation include oncogenes, such as RAS [19] and Phospholipase C epsilon 1 (PLCE1)[20], whose hypomethylation indicates a promising therapeutic target. However, some tumor suppressor genes show hypermethylation in tumors, resulting in decreased gene expression and induction of tumorigenesis and development; these hypermethylated tumor suppressors include Ras association domain family member 1 (RASSF1) methylation in bladder cancer, O-6-methylguanine-DNA methyltransferase (MGMT) methylation in non-small-cell lung cancer and retinoic acid receptor beta (RARβ) methylation in prostate cancer [21]. RASSF1 has been characterized as a tumor suppressor gene because it has the potential to inhibit cell proliferation, control the cell cycle and promote apoptosis. The MGMT protein has shown DNA repair ability because it prevents the formation of DNA adducts of carcinogens. RARβ has been shown to function as a tumor suppressor gene, and inactivation of its promoter via DNA methylation is associated with lung, breast and prostate cancer. In addition to the hypermethylation of the RARβ gene, it was recently reported that Kruppel-

gene, KLF4 has a hypermethylated promoter sequence, and its gene expression is downregulated in prostate cancer, which promotes tumor proliferation and drug resistance. Downregulation of KLF4 or hypermethylation of KLF4 gene promoters can be considered markers for predicting the early recurrence of cancer[22, 23]. GSTP1 is functionally inactivated by promoter methylation, which increases the sensitivity to oxidative stress and increases the risk of prostate cancer progression [24]. Studies have shown that GSTP1 can be used as a biomarker of fluid biopsy in prostate cancer, but the results show that it has high specificity and low sensitivity [25]. A post hoc analysis of data from a Phase 3 trial showed that the detection of serum free methylated GSTP1 (mGSTP1) DNA can predict the outcome of chemotherapy in metastatic prostate cancer and may guide treatment decisions in the clinic [26]. In addition, methylation of other genes, including olfactory 4 (OLFM4)[27], MTSS1[7] and myeloid ecotropic viral insertion site 2 (MEIS2)[28], also plays an important role in the occurrence and development of prostate cancer. The methylation pattern of prostate cancer can be used as a potential reasonable choice for disease classification and prognosis evaluation. At present, DNA methylation sites related to the prognosis of prostate cancer have been found, including GPR7 (known as neuropeptides B and W receptor 1, NPBWR1), ABHD9 (known as epoxide hydrolase 3, EPHX3), an expressed sequence tag on chromosome 3 (Chr3-EST) [29], paired like homeodomain 2 (PITX2)[30] and zinc finger protein 154 (ZNF154)[31]. The clinical significance of these genes with different methylation patterns in tumor classification, survival and prognosis has yet to be examined in a large prostate cancer dataset. In this study, we tried to explore a classification method that integrates several DNA methylation markers to help clinicians evaluate the therapeutic effect and prognosis of prostate cancer patients and choose treatment strategies.

In general, we used a large dataset from the TCGA and GEO databases to screen methylation sites at the prognostic level by means of bioinformatics. We used the LASSO algorithm to eliminate the methylation sites with small differences; we retained the 21 methylation sites with the most significant differences, and used their corresponding gene expression data to construct a recurrence prediction model for patients with prostate cancer. We observed methylation models in terms of immunity, pathway, and genomic variation.

Flammiger et al [17] detected Treg cells in prostate cancer tissues by FOXP3 immunohistochemistry and found that prostate-specific antigen (PSA) recurrence-free survival was reduced in patients with high levels of Treg cells, which was associated with more advanced tumor stages and a higher Ki67 labeling index. Erlandsson et al [16] found that the higher the degree of tumor infiltration of M2 macrophages was, the greater the risk of death in patients with prostate cancer; the high degree of infiltration of M2 macrophages and other inhibitory cells (such as Tregs) may support the immunosuppressive microenvironment. The role of B cells in tumors is not clear. Many studies have reported that tumor-infiltrating B cells may promote tumor progression. However, other studies have shown that tumor infiltrating B cells may be associated with a good prognosis and an enhanced response to chemotherapy[32]. WooJR et al[18] found that the degree of B lymphocyte infiltration in prostate cancer tissues was higher than that in adjacent tissues, and the degree of B lymphocyte infiltration in patients with a high risk of recurrence was higher than that in patients with a low risk of recurrence. The role of B cell immunity in prostate cancer is unclear, which indicates opportunities for research and potential treatment.

According to the gene enrichment analysis, we determined that the related metastatic gene sets of prostate cancer are enriched in cluster 3 subtypes, and it can be inferred that the genes that distinguish subtypes play an important role in the metastasis of prostate cancer. In addition, our results suggest that the SUMOylation of DNA methyltransferase may play a vital role in tumor metastasis.

Our results show that patients in cluster 3 have a higher burden of somatic mutations in TP53, TTN and KMT2D. p53 is a tumor suppressor that responds to DNA damage[33] by regulating the cell cycle checkpoint pathway and tends to mutate in advanced, recurrent and metastatic prostate cancer [34]. Missense mutation of titin (TTN) is correlated with favorable prognosis in lung squamous cell carcinoma [35] In multiple tumor cohorts, the progression-free survival time or overall survival time of patients with TTN mutations is higher than that of patients with wild-type TTN, which is not consistent with our results. The role of TTN in prostate cancer needs to be further determined. KMT2D is a histone H3 lysine 4 (H3K4) methyltransferase that inhibits the growth and metastasis of bladder cancer cells by regulating tumor suppressor genes such as TP53[36]. Our study provides new ideas regarding the potential mechanism by which the high TP53, TTN and KMT2D mutation burden in cluster 3 subtypes may promote the development of prostate cancer through increased DNA damage and dysregulation of tumor suppressor genes.

Androgen receptors are members of the nuclear receptor superfamily that play a key role in prostate cancer. The upregulation of AR may be the cause of the development of prostate cancer into castration-resistant prostate cancer and the related poor prognosis [37]. Studies show that metadherin (MTDH) plays a role in the progression of various tumors [38]. In prostate cancer, MTDH acts as an oncogene, regulating the growth and metastasis of prostate cancer cells [39].

Finally, the effectiveness and universality of the methylation model were verified and evaluated. Our study shows that classification based on DNA methylation is an independent factor for predicting tumor recurrence in patients with prostate cancer. Our model can provide clinicians with a judgment of prognosis in patients with prostate cancer and provide them with guidance for choosing treatment strategies. However, our research has some limitations. Our methylation model was established using data from prostate cancer. Further studies are needed to determine whether it is suitable for different subtypes of prostate cancer, such as castration-resistant prostate cancer and metastatic prostate cancer.

Conclusions

We conclude that classification based on DNA methylation is an independent factor for predicting tumor recurrence in patients with prostate cancer. Our model can contribute to new biomarker identification, drug targeting and molecular subtype classification in patients with prostate cancer.

Abbreviations

AN: Adjacent normal;

AR: Androgen receptor;

ATP1B2: ATPase Na⁺/K⁺ transporting subunit beta 2;

Chr3-EST: an expressed sequence tag on chromosome 3

CNV: Copy number variation;

DFS: Disease-free survival;

DOCK2: Dedicator of cytokinesis 2;

FAM92B: Family with sequence similarity 92, member B;

GEO: Gene Expression Omnibus

GSTP1: Glutathione S-transferase π ;

KLF4: Kruppel-like factor 4;

KMT2D: Lysine methyltransferase 2D;

LAPTM4B: Lysosomal protein transmembrane 4 beta;

LASSO: Least Absolute Shrinkage and Selection Operator;

MEIS2: Myeloid Ecotropic viral Insertion Site 2;

MGMT: O-6-methylguanine-DNA methyltransferase;

MTDH: Metadherin;

MTSS1: Metastasis suppressor 1;

NPBWR1: Neuropeptides B and W receptor 1;

OLFM4: Olfactory 4;

OS: Overall survival;

PC: Prostate cancer;

PITX2: Paired like homeodomain 2;

PLCE1: Phospholipase C epsilon 1;

RARB: Retinoic acid receptor beta;

RASSF1: Ras association domain family member 1;

TCGA: The Cancer Genome Atlas

TMB: Tumor mutational burden;

TP53: Tumor protein p53;

TTN: Titin;

WES: Whole Exome Sequencing;

ZNF154: zinc Finger protein 154;

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The dataset used and/or analyzed during the current study are available from the TCGA network, and from Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>, GEO accessions: GSE2612).

Competing interests

The authors declare that they have no competing interests.

Funding

Project was supported by the Shanghai Committee of Science and Technology (Grant No.19441909400).

Authors' contributions

The conception and design of the study was taken over by SJX, YS, LC and DSP. MLS, LC, DL, CLX, FTC and JL were involved in data acquisition, statistical analyses, and data interpretation. The manuscript was

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

drafted by DSP with the help of LC. The critical revision of the manuscript was performed by SJX and YS. All authors have read and approved the manuscript.

Acknowledgements

Not applicable

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018, **68**:394-424.
2. Rawla P: **Epidemiology of Prostate Cancer.** *World J Oncol* 2019, **10**:63-89.
3. **Global cancer observatory: cancer tomorrow.** [<https://gco.iarc.fr/tomorrow>]
4. Feinberg AP, Tycko B: **The history of cancer epigenetics.** *Nat Rev Cancer* 2004, **4**:143-153.
5. Wu Y, Sarkissyan M, Vadgama JV: **Epigenetics in breast and prostate cancer.** *Methods Mol Biol* 2015, **1238**:425-466.
6. Mian OY, Khattab MH, Hedayati M, Coulter J, Abubaker-Sharif B, Schwaninger JM, Veeraswamy RK, Brooks JD, Hopkins L, Shinohara DB, et al: **GSTP1 Loss results in accumulation of oxidative DNA base damage and promotes prostate cancer cell survival following exposure to protracted oxidative stress.** *Prostate* 2016, **76**:199-206.
7. Chen J, Huang L, Zhu Q, Wang Z, Tang Z: **MTSS1 hypermethylation is associated with prostate cancer progression.** *J Cell Physiol* 2020, **235**:2687-2697.
8. Bjerre MT, Strand SH, Norgaard M, Kristensen H, Rasmussen AK, Mortensen MM, Fredsoe J, Mouritzen P, Ulhoi B, Orntoft T, et al: **Aberrant DOCK2, GRASP, HIF3A and PKFP Hypermethylation has Potential as a Prognostic Biomarker for Prostate Cancer.** *Int J Mol Sci* 2019, **20**.
9. Liu L, Kron KJ, Pethe VV, Demetrashvili N, Nesbitt ME, Trachtenberg J, Ozcelik H, Fleshner NE, Briollais L, van der Kwast TH, Bapat B: **Association of tissue promoter methylation levels of APC, TGFbeta2, HOXD3 and RASSF1A with prostate cancer progression.** *Int J Cancer* 2011, **129**:2454-2462.
10. Gao T, He B, Pan Y, Li R, Xu Y, Chen L, Nie Z, Gu L, Wang S: **The association of retinoic acid receptor beta2(RARbeta2) methylation status and prostate cancer risk: a systematic review and meta-analysis.** *PLoS One* 2013, **8**:e62950.
11. **the Cancer Genome Atlas** [<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>]
12. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R: **Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.** *Epigenetics* 2013, **8**:203-209.
13. Liu Q, Liu C: **A Novel Locally Linear KNN Method With Applications to Visual Recognition.** *IEEE Trans*

14. Li G, Xu W, Zhang L, Liu T, Jin G, Song J, Wu J, Wang Y, Chen W, Zhang C, et al: **Development and validation of a CIMP-associated prognostic model for hepatocellular carcinoma.** *EBioMedicine* 2019, **47**:128-141.
15. Tibshirani R: **The lasso method for variable selection in the Cox model.** *Stat Med* 1997, **16**:385-395.
16. Erlandsson A, Carlsson J, Lundholm M, Falt A, Andersson SO, Andren O, Davidsson S: **M2 macrophages and regulatory T cells in lethal prostate cancer.** *Prostate* 2019, **79**:363-369.
17. Flammiger A, Weisbach L, Huland H, Tennstedt P, Simon R, Minner S, Bokemeyer C, Sauter G, Schlomm T, Trepel M: **High tissue density of FOXP3+ T cells is associated with clinical outcome in prostate cancer.** *Eur J Cancer* 2013, **49**:1273-1279.
18. Woo JR, Liss MA, Muldong MT, Palazzi K, Strasner A, Ammirante M, Varki N, Shabaik A, Howell S, Kane CJ, et al: **Tumor infiltrating B-cells are increased in prostate cancer tissue.** *J Transl Med* 2014, **12**:30.
19. Maryam M, Idrees M: **Study of promoter hypomethylation profiles of RAS oncogenes in hepatocellular carcinoma derived from hepatitis C virus genotype 3a in Pakistani population.** *J Med Virol* 2018, **90**:1516-1523.
20. Chen Y, Wang D, Peng H, Chen X, Han X, Yu J, Wang W, Liang L, Liu Z, Zheng Y, et al: **Epigenetically upregulated oncoprotein PLCE1 drives esophageal carcinoma angiogenesis and proliferation via activating the PI-PLCepsilon-NF-kappaB signaling pathway and VEGF-C/ Bcl-2 expression.** *Mol Cancer* 2019, **18**:1.
21. Bouras E, Karakioulaki M, Bougioukas KI, Aivaliotis M, Tzimagiorgis G, Chourdakis M: **Gene promoter methylation and cancer: An umbrella review.** *Gene* 2019, **710**:333-340.
22. Jiang Z, Zhang Y, Chen X, Wu P, Chen D: **Long non-coding RNA LINC00673 silencing inhibits proliferation and drug resistance of prostate cancer cells via decreasing KLF4 promoter methylation.** *J Cell Mol Med* 2020, **24**:1878-1892.
23. Li H, Wang J, Xiao W, Xia D, Lang B, Wang T, Guo X, Hu Z, Ye Z, Xu H: **Epigenetic inactivation of KLF4 is associated with urothelial cancer progression and early recurrence.** *J Urol* 2014, **191**:493-501.
24. Zhang L, Meng X, Pan C, Qu F, Gan W, Xiang Z, Han X, Li D: **piR-31470 epigenetically suppresses the expression of glutathione S-transferase pi 1 in prostate cancer via DNA methylation.** *Cell Signal* 2020, **67**:109501.
25. Gurioli G, Martignano F, Salvi S, Costantini M, Gunelli R, Casadio V: **GSTP1 methylation in cancer: a liquid biopsy biomarker?** *Clin Chem Lab Med* 2018, **56**:702-717.
26. Mahon KL, Qu W, Lin HM, Spielman C, Cain D, Jacobs C, Stockler MR, Higano CS, de Bono JS, Chi KN, et al: **Serum Free Methylated Glutathione S-transferase 1 DNA Levels, Survival, and Response to Docetaxel in Metastatic, Castration-resistant Prostate Cancer: Post Hoc Analyses of Data from a Phase 3 Trial.** *Eur Urol* 2019, **76**:306-312.
27. Li H, Kim C, Liu W, Zhu J, Chin K, Rodriguez-Canales J, Rodgers GP: **Olfactomedin 4 downregulation is associated with tumor initiation, growth and progression in human prostate cancer.** *Int J Cancer* 2020, **146**:1346-1358.

28. Norgaard M, Haldrup C, Bjerre MT, Hoyer S, Ulhoi B, Borre M, Sorensen KD: **Epigenetic silencing of MEIS2 in prostate cancer recurrence.** *Clin Epigenetics* 2019, **11**:147.
29. Cottrell S, Jung K, Kristiansen G, Eltze E, Semjonow A, Ittmann M, Hartmann A, Stamey T, Haefliger C, Weiss G: **Discovery and validation of 3 novel DNA methylation markers of prostate cancer prognosis.** *J Urol* 2007, **177**:1753-1758.
30. Jiang Q, Xie M, He M, Yan F, Chen M, Xu S, Zhang X, Shen P: **PITX2 methylation: a novel and effective biomarker for monitoring biochemical recurrence risk of prostate cancer.** *Medicine (Baltimore)* 2019, **98**:e13820.
31. Zhang W, Shu P, Wang S, Song J, Liu K, Wang C, Ran L: **ZNF154 is a promising diagnosis biomarker and predicts biochemical recurrence in prostate cancer.** *Gene* 2018, **675**:136-143.
32. Flynn NJ, Somasundaram R, Arnold KM, Sims-Mourtada J: **The Multifaceted Roles of B Cells in Solid Tumors: Emerging Treatment Opportunities.** *Target Oncol* 2017, **12**:139-152.
33. Ko LJ, Prives C: **p53: puzzle and paradigm.** *Genes Dev* 1996, **10**:1054-1072.
34. Voeller HJ, Sugars LY, Pretlow T, Gelmann EP: **p53 oncogene mutations in human prostate cancer specimens.** *J Urol* 1994, **151**:492-495.
35. Cheng X, Yin H, Fu J, Chen C, An J, Guan J, Duan R, Li H, Shen H: **Aggregate analysis based on TCGA: TTN missense mutation correlates with favorable prognosis in lung squamous cell carcinoma.** *J Cancer Res Clin Oncol* 2019, **145**:1027-1035.
36. Sun P, Wu T, Sun X, Cui Z, Zhang H, Xia Q, Zhang D: **KMT2D inhibits the growth and metastasis of bladder Cancer cells by maintaining the tumor suppressor genes.** *Biomed Pharmacother* 2019, **115**:108924.
37. Fujita K, Nonomura N: **Role of Androgen Receptor in Prostate Cancer: A Review.** *World J Mens Health* 2019, **37**:288-295.
38. Hu G, Wei Y, Kang Y: **The multifaceted role of MTDH/AEG-1 in cancer progression.** *Clin Cancer Res* 2009, **15**:5615-5620.
39. Pan D, Jia Z, Li W, Dou Z: **The targeting of MTDH by miR1455p or miR1453p is associated with prognosis and regulates the growth and metastasis of prostate cancer cells.** *Int J Oncol* 2019, **54**:1955-1968.

Figures

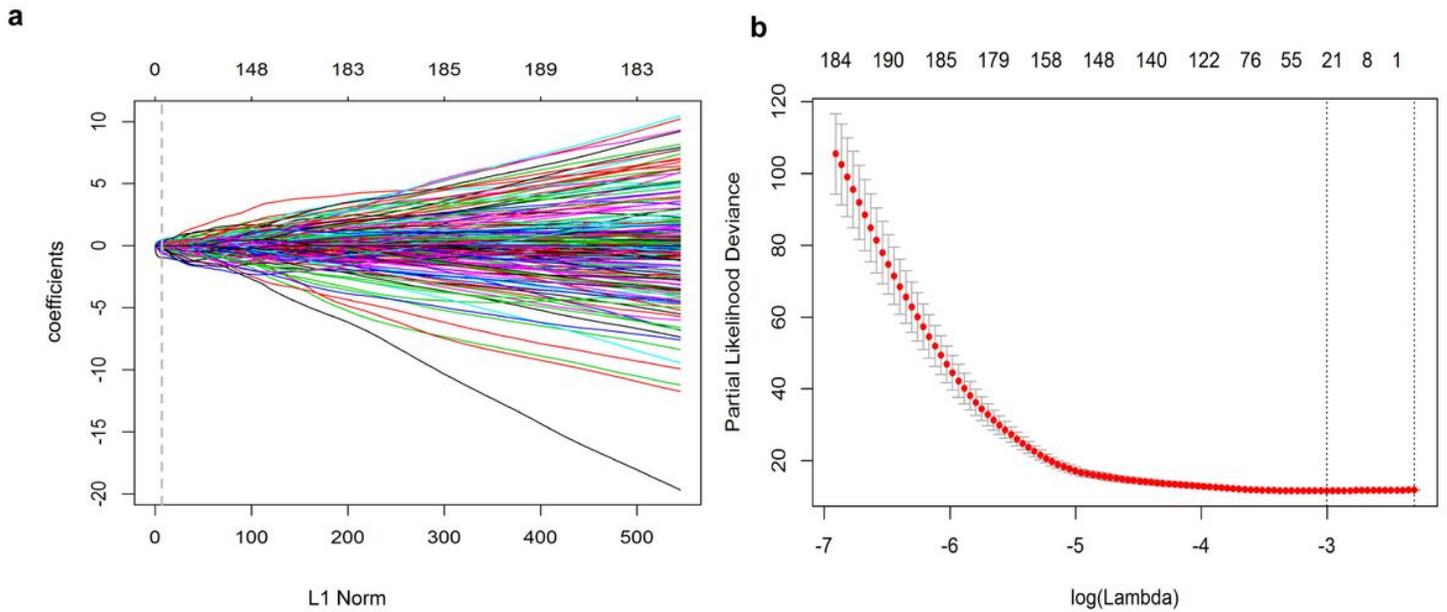


Figure 1

a. The changing trajectory of each independent variable. The horizontal axis represents the log value of the independent variable lambda, and the vertical axis represents the coefficient of the independent variable. b. Confidence intervals for each lambda. The dotted line on the left represents the best penalty coefficient.

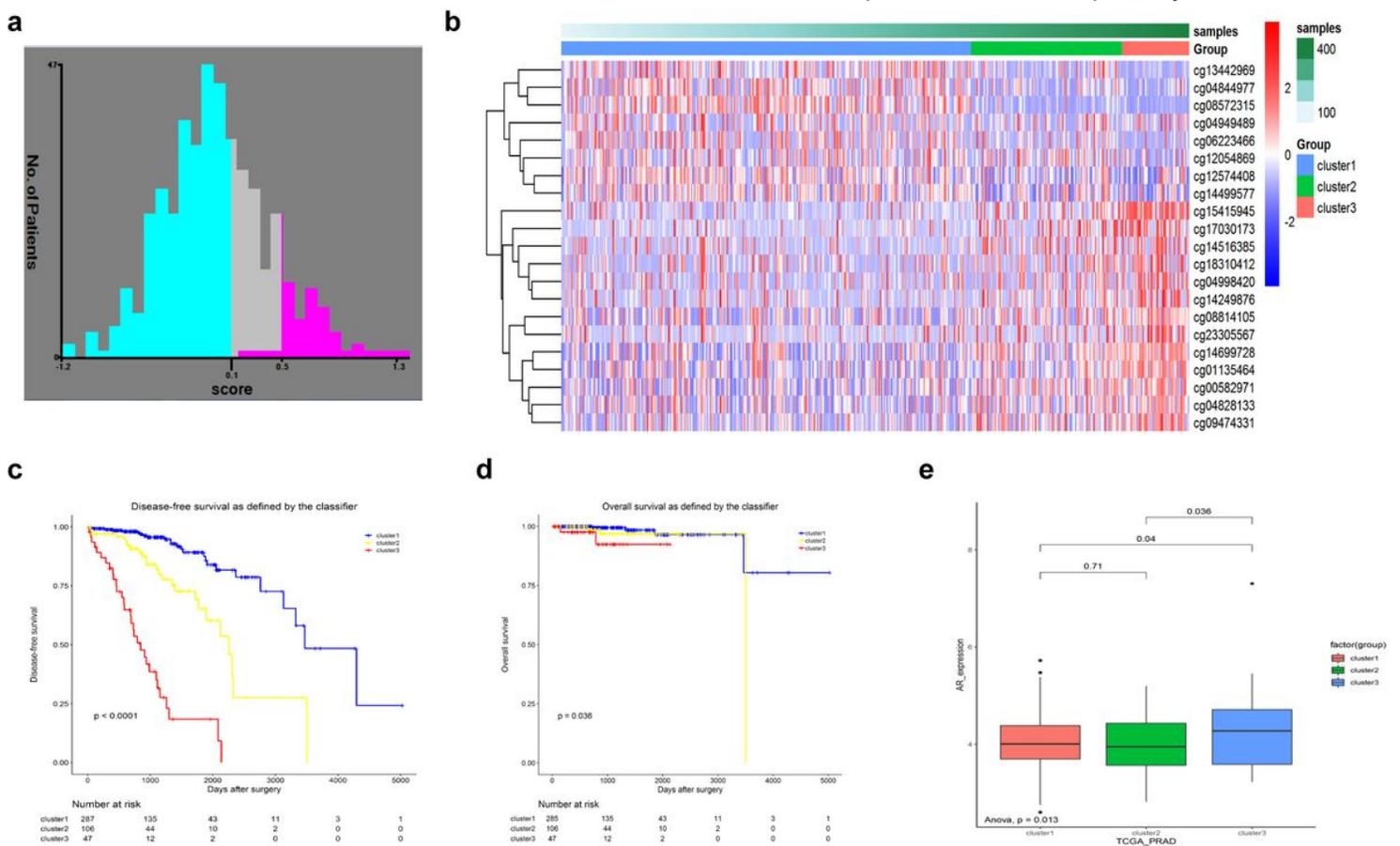


Figure 2

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

a. The cutoff plot classifies patients into three groups: the low-risk group, medium-risk group and high-risk group. b. The differential methylation heat maps of the three groups. c. The prognostic difference (disease-free survival, DFS) after classification prediction in the training set. d. The prognostic difference (overall survival, OS) after classification prediction in the training set. e. Expression of AR in three groups of methylation subtypes, $p=0.013$.

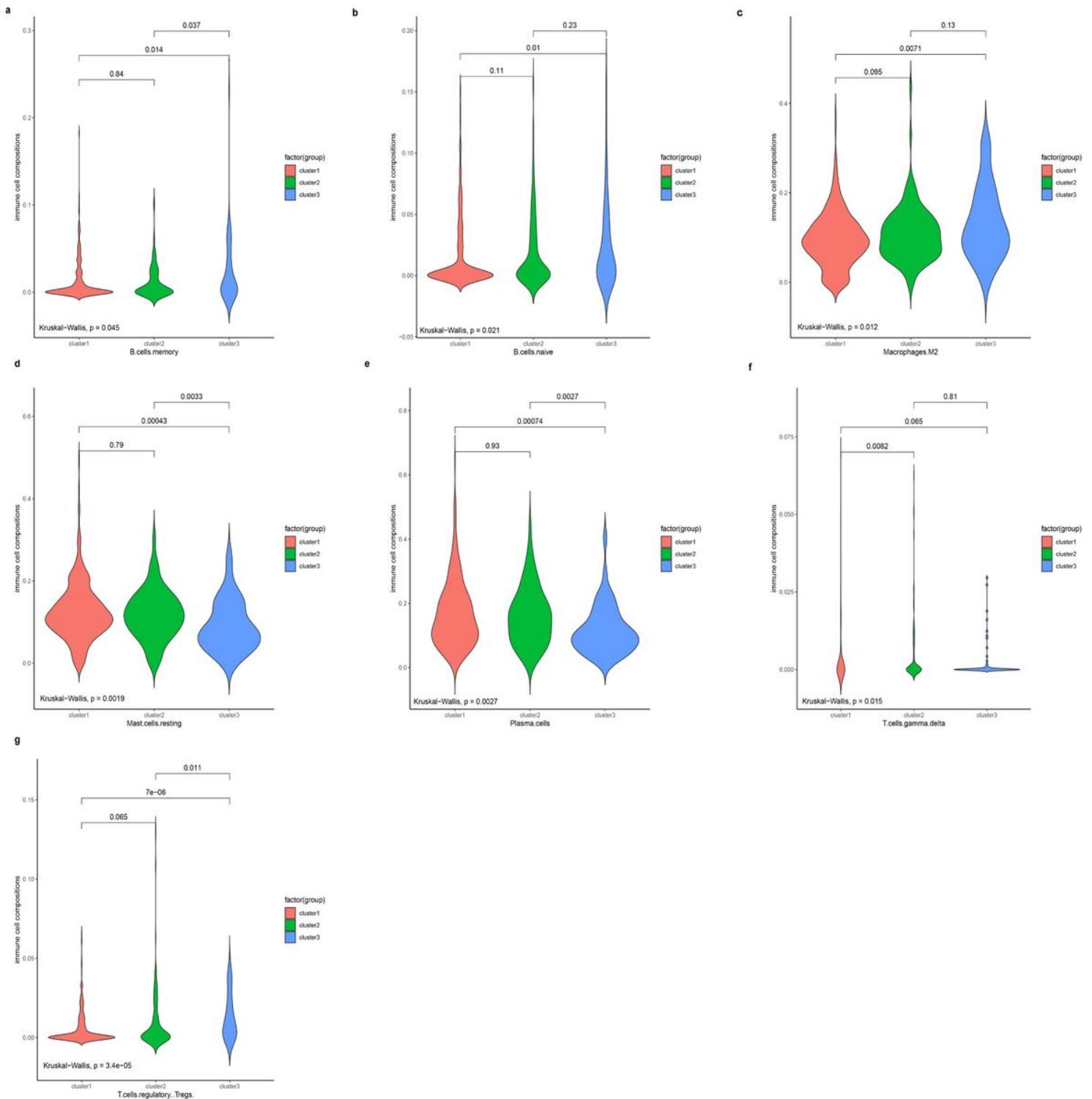


Figure 3

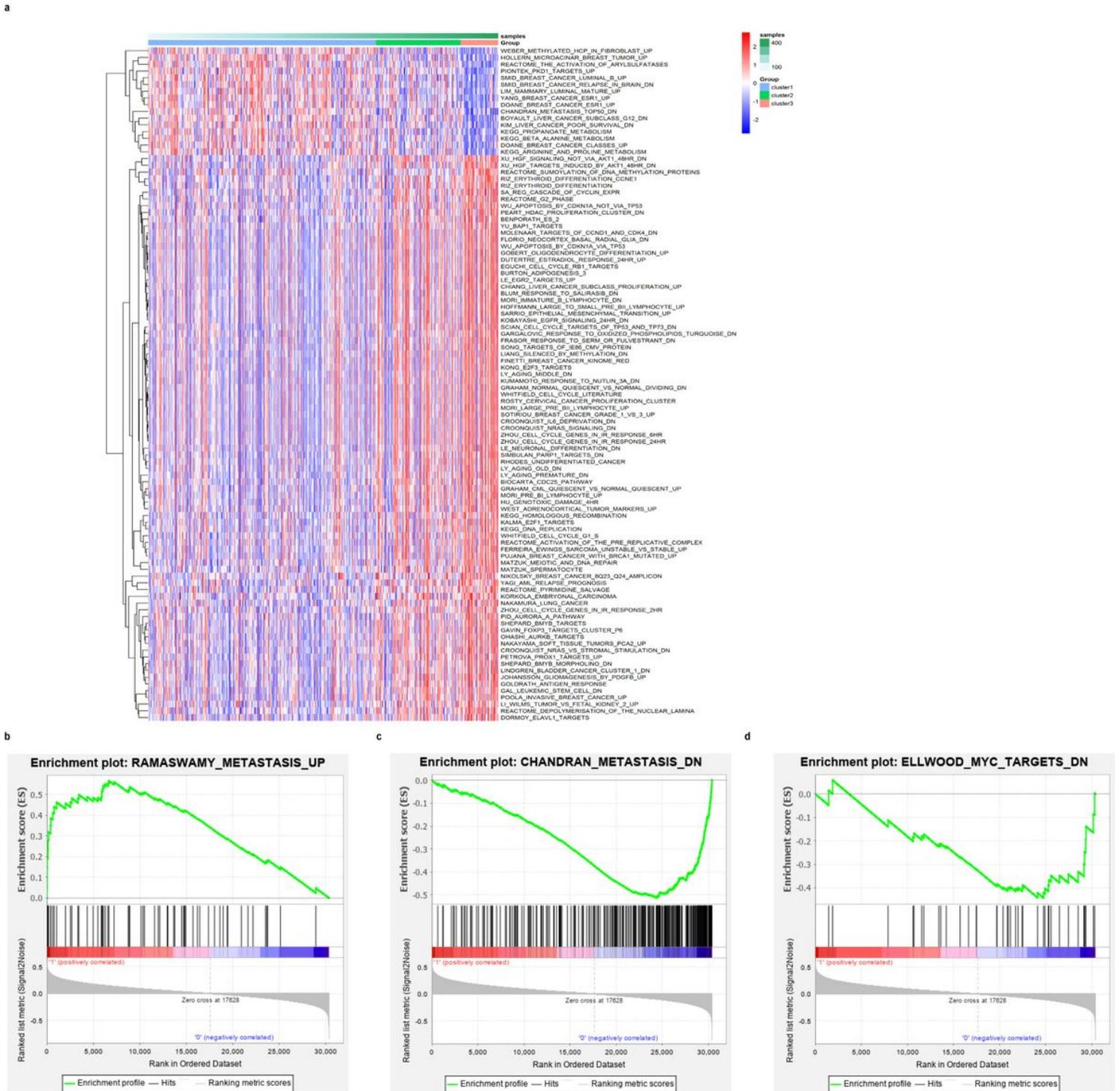


Figure 4

a. A heat map of the gene set. b-d. Significant enrichment in cluster 3 compared with cluster 1 and cluster 2 in prostate cancer. "RAMASWAMY_METASTASIS_UP" is a set of genes Up-regulated in metastatic vs primary solid tumors. "CHANDRAN_METASTASIS_DN" is a set of Genes down-regulated in metastatic tumors from the whole panel of patients with prostate cancer. "ELLWOOD_MYC_TARGETS_DN" is a set of Genes down-regulated in transgenic mice expressing human MYC in prostate.

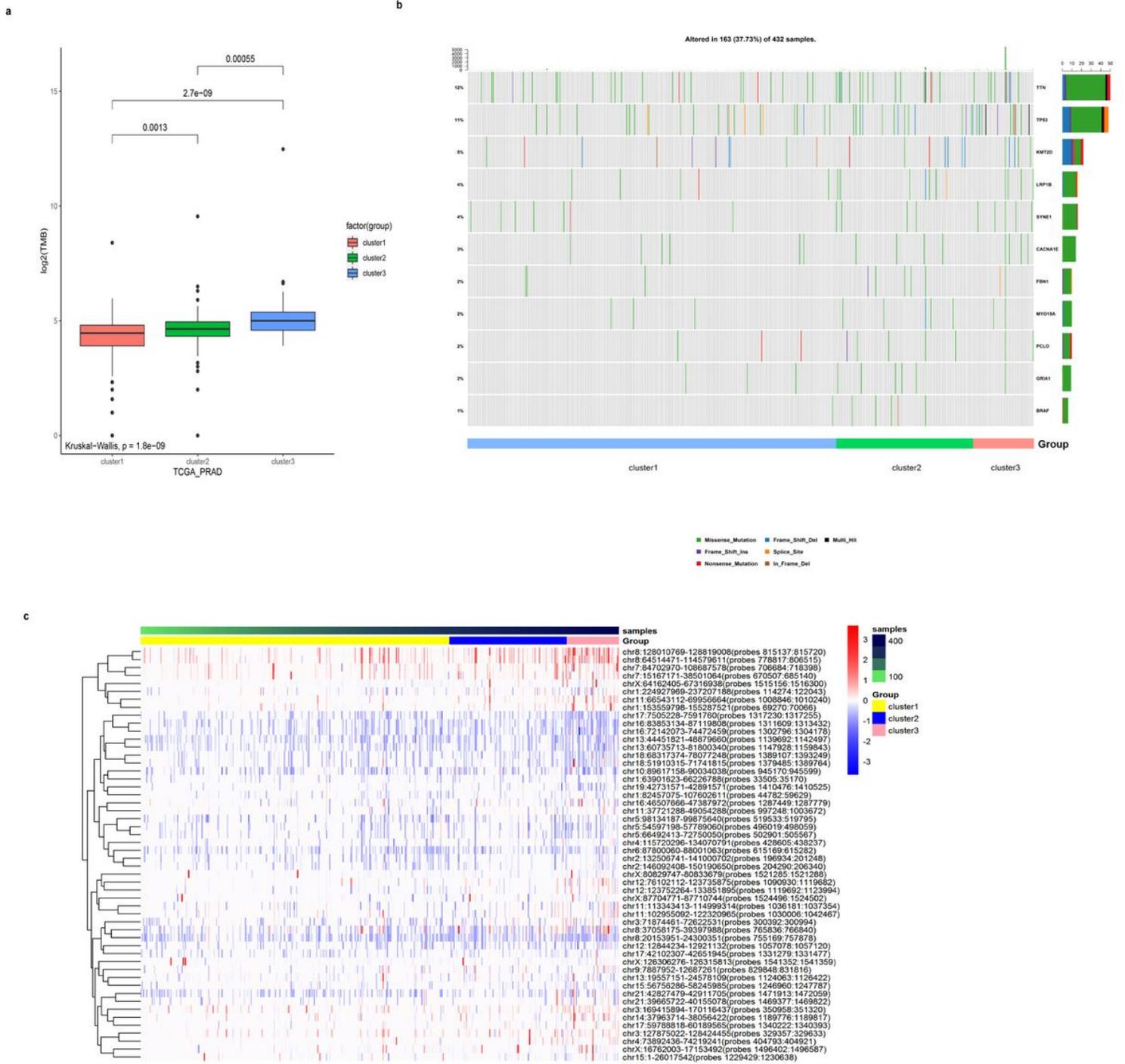


Figure 5

a. Differences in TMB among the three subtypes, $p < 0.001$. b. TMB of different genes in different subtypes. c. Gene annotation of CNV changes of different subtypes in prostate cancer.

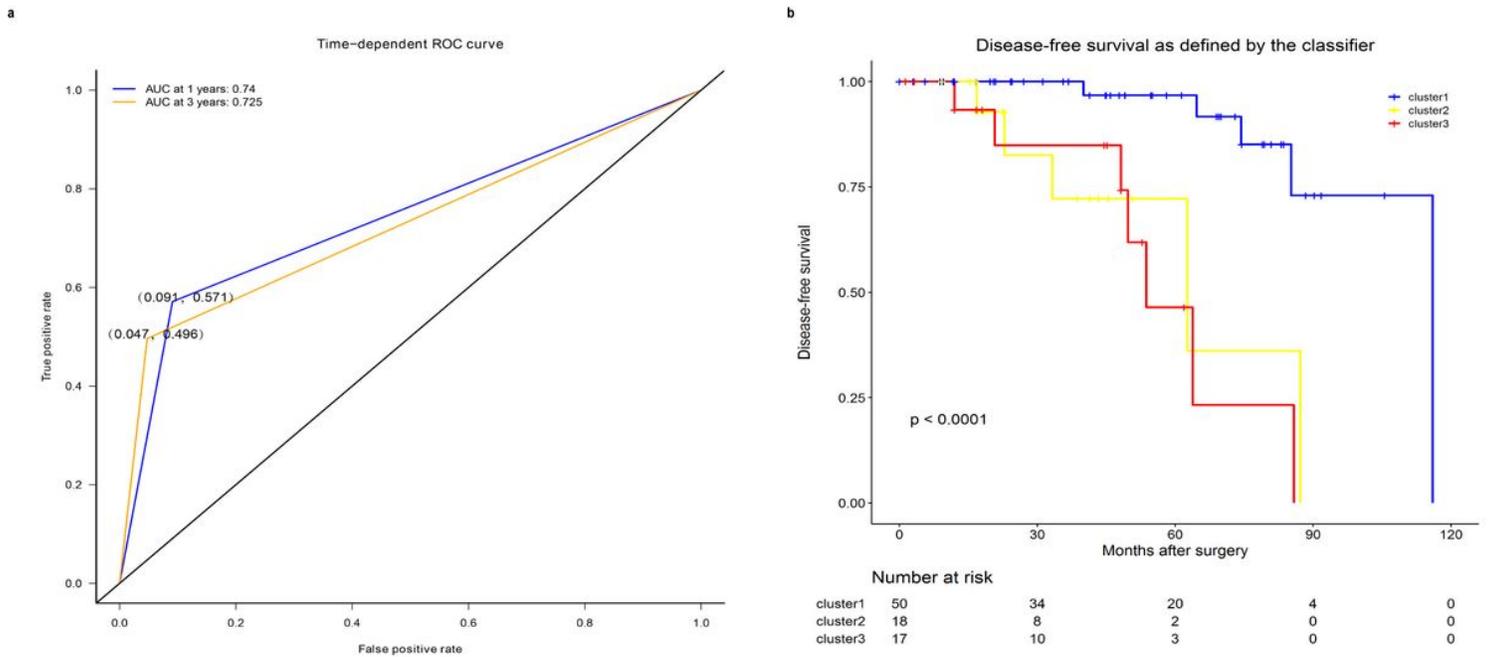


Figure 6

a. ROC curves showing the sensitivity and specificity of the methylation model in predicting recurrence in patients. b. The prognostic difference (DFS) after the classification prediction in the validation set.

Supplementary Files

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

- [TableS3.xlsx](#)
- [TableS2.xlsx](#)
- [TableS1.xlsx](#)
- [FigS2.pdf](#)
- [FigS1.pdf](#)