

An 8-gene DNA methylation signature predicts the recurrence risk of cervical cancer

Journal of International Medical Research
49(5) 1–14

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03000605211018443

journals.sagepub.com/home/imr



Jing-Hang Ma^{1,2}, Yu Huang¹, Lu-Yao Liu¹ and Zhen Feng¹ 

Abstract

Objective: This study examined the predictive utility of DNA methylation for cervical cancer recurrence.

Methods: DNA methylation and RNA expression data for patients with cervical cancer were downloaded from The Cancer Genome Atlas. Differentially methylated genes (DMGs) and differentially expressed genes were screened and extracted *via* correlation analysis. A support vector machine (SVM)-based recurrence prediction model was established using the selected DMGs. Cox regression analysis and receiver operating characteristic curve analysis were used for self-evaluation. The Gene Expression Omnibus (GEO) database was applied for external validation. Functional enrichment was determined using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses.

Results: An eight-gene DNA methylation signature identified patients with a high risk of recurrence (area under the curve = 0.833). The SVM score was an independent risk factor for recurrence (hazard ratio [HR] = 0.418; 95% confidence interval [CI] = 0.26–0.67). The independent GEO database analysis further supported the result.

Conclusion: An eight-gene DNA methylation signature predictive of cervical cancer recurrence was identified in this study, and this signature may help identify patients at high risk of recurrence and improve clinical treatment.

¹First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

²Department of Gynecology, First Affiliated Hospital, Wenzhou Medical University, Wenzhou, China

Corresponding author:

Zhen Feng, First Affiliated Hospital of Wenzhou Medical University, Fanhai West Road, Wenzhou 325000, China. Email: fengzhen@wmu.edu.cn



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative

Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Keywords

DNA methylation, cancer recurrence, prediction model, support vector machine, differentially expressed gene, cervical cancer

Date received: 25 January 2021; accepted: 26 April 2021

Introduction

Cervical cancer is the leading cause of cancer-related death in women, being responsible for more than 500,000 diagnoses and 260,000 deaths each year worldwide.¹ Surgery is the main treatment for early-stage cervical cancer. For patients at high risk of recurrence, radiotherapy alone or in combination with chemotherapy is also needed.² It has been reported that the recurrence rate of cervical cancer is 20% to 30%, and this rate increases with disease progression. Patients with recurrent disease always have poor prognoses, and therapeutic methods with sufficient response rates are limited.^{3–5}

At present, staging systems (FIGO or TNM staging) together with the “Sedlis criteria” and other risk factors are generally used to identify patients with cervical cancer at high risk of recurrence according to clinical and pathological characteristics to guide adjuvant treatment.^{6,7} These assessments of recurrence risk undoubtedly provide predictive value; however, they are less accurate for predicting cancer recurrence than genetic, epigenetic, and molecular biomarkers. Therefore, it is necessary to identify biomarkers for predicting recurrence risk and determining treatment strategies to improve patients’ prognoses.^{8,9}

DNA methylation regulates gene transcription without changing the DNA sequence. Abnormal hypermethylation of tumor suppressor genes and hypomethylation of oncogenes are vital during the early tumorigenesis process,^{10,11} and related

epigenetic drugs have proven effective in the treatment of cervical cancer.¹² Compared with the complex variability in RNA expression between individual patients,¹³ DNA methylation is relatively stable, and pervasive modification across tumor types is associated with the pathogenesis and progression of cancer and is related to patient prognosis and outcome.¹³ Researchers have explored the relationships between DNA methylation and cervical cancer, including the biological function of genes,^{14–16} methylation-based tumor classification,^{17,18} and radiotherapy-related predictors.¹⁹ However, few studies have sought to develop recurrence prediction models for DNA methylation in cervical cancer, which is of clinical significance to help clinicians identify high-risk patients and implement active treatment.

In this study, we used DNA methylation and related RNA expression datasets for patients with cervical cancer from The Cancer Genome Atlas (TCGA) to identify an epigenetic signature for recurrence risk prediction.

Materials and methods

Sample selection and data processing

DNA methylation datasets, RNA expression datasets, and clinical information for patients with cervical cancer used for modeling were downloaded from TCGA (<https://portal.gdc.cancer.gov/>)²⁰ using the FireBrowse Data Portal (<http://firebrowse>).

org/).²¹ Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA) sequencing technology was used to obtain 312 DNA methylation profiles, and HiSeq2000 (Illumina) was used to analyze 307 mRNA-seq profiles.

All samples in TCGA were collected from patients with primary cervical tumors, with tumor purity ranging from 22% to 96%.²⁰ We excluded samples from patients with missing survival information and included samples according to the recurrence status and disease-free survival.

All data used in the study were obtained from public databases, which are open source. Thus, the requirement for ethical approval was waived.

Differential methylation analysis and differential expression analysis

DNA methylation datasets and RNA expression datasets from TCGA were analyzed using R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria). Genes with differences in DNA methylation levels (β values) significant at $P < 0.01$ with an absolute log fold change exceeding 0.025 were considered differentially methylated genes (DMGs). Genes with differences in RNA expression significant at $P < 0.01$ with absolute log fold change exceeding 0.4 were considered differentially expressed genes (DEGs). The log fold change thresholds were set to identify genes proximately ranked in the top 1%. DMGs overlapping with DEGs were extracted for the correlation analysis. We used the *lm* function in R to fit the linear model. Genes with a significant negative correlation between DMGs and DEGs were further chosen to establish the support vector machine (SVM)-based prediction model.

SVM model

SVM is a machine-learning algorithm that offers a direct approach to binary classification.²² The SVM considers data as points in a high-dimensional space and finds an optimal hyperplane to separate data into classes.

The eight selected DMGs from patients in the TCGA database were denoted as 135 pairs $(x_1, y_1), (x_2, y_2), \dots, (x_{135}, y_{135})$. $x_i \in \mathbb{R}^8$ denotes the eight-dimensional data that represent the methylation levels of eight DMGs in the i th sample, and $y_i \in \{-1, 1\}$ represents the labels of classification. Each pair (x_i, y_i) was mapped onto data points in the eight-dimensional space and used to train the SVM-based prediction model without cross-validation. The model found the optimal linear hyperplane $P0\{x: \varphi(x) = 0\}$ in the feature space, where φ was the linear transformation. Points satisfying $\{x: \varphi(x) > 0\}$ or $\{x: \varphi(x) < 0\}$ belonged to the class $y = 1$ or $y = -1$, respectively (the graphical illustration is presented in Figure 3b).

We extracted the φ function along with its coefficients from the linear SVM model and generated the SVM score to predict the classification. The SVM score was calculated using the β values of the eight DMGs and their coefficients in the trained SVM model. A constant bias was also introduced. A sample was classified into the low-risk or high-risk class if the SVM score was > 0 or ≤ 0 , respectively. In this work, we used the *e1071* package in R,²³ which contained the *svm* function.

Recurrent signature validation

The confusion matrix was constructed to further confirm the prediction of the SVM classifier model. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) were used to assess the fit of the SVM-based prediction model.²⁴ A *t*-test and Fisher's exact test

were used to identify clinical features affecting recurrence. The terms significant at $P < 0.05$ were retained for further multivariate Cox regression analysis.

Data visualization using t-distributed stochastic neighbor embedding (t-SNE)

The t-SNE algorithm,²⁵ a nonlinear dimensionality reduction algorithm, was used to maintain the topological structure in the high dimension and map each datum onto several different, but related, low-dimensional manifolds because an SVM plot cannot visualize the eight-dimensional data of the eight DMGs.

We used the Rtsne package in R.²⁶ The parameter “perplexity,” which balanced attention between local and global aspects of the data and which was an estimate of the approximately 30 closest neighbors each data point had, was set to 30. The number of iterations was 10,000.

External validation of the Gene Expression Omnibus (GEO) database

We searched cervical cancer datasets containing both DNA methylation and survival information across the GEO database, and the only GSE30759 dataset (Illumina Infinium 27k Human Methylation BeadChip for DNA methylation profiles) with 63 samples of primary cervical cancer tissue satisfied such conditions. We preprocessed the samples as described for TCGA datasets. Nonrecurrence was indicated by recurrence-free survival of at least 2 years, and all other patients were considered to have recurrent cancer.

The recurrence-related methylated genes in the GEO dataset overlapped with the eight reference genes obtained from the training model using TCGA dataset. Three genes, namely *LOC100132215*, *TIGIT*, and *ZFYVE21*, in the GEO dataset were found to not have probes. We calculated the

average DNA methylation levels of these three genes in TCGA dataset and subtracted them from the SVM score threshold according to equation 2, eventually generating the reduced SVM score threshold.

We analyzed the distribution of the methylation levels for the other five genes in both datasets. We found four genes with similar distributions in both datasets. We then normalized the levels in the GEO dataset to those in TCGA datasets to avoid bias. Normalization of the methylation value of genes from the i^{th} sample in the GEO dataset was performed as follows:

$$\beta_{Gene,GEO}^i = \frac{\langle \beta_{Gene,TCGA} \rangle}{\langle \beta_{Gene,GEO} \rangle} \times \beta_{Gene,GEO}^i \quad (1)$$

Of the five genes, *MPHOSPH10* was excluded because of the large differences in its methylation between the two datasets. Similarly, we subtracted the average DNA methylation level of *MPHOSPH10* in TCGA dataset to further update the reduced SVM score threshold.

Functional enrichment analysis

Gene Ontology (GO) enrichment analysis²⁷ was performed to identify the function of DMGs, including biological process (BP), cellular component (CC), and molecular function (MF). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed to identify KEGG pathways. The Database for Annotation, Visualization and Integrated Discovery was used to perform GO and KEGG enrichment analyses (<https://david.ncifcrf.gov/>).²⁸ The enrichment functions and pathways were considered significant at $P < 0.05$.

Results

The study procedures are presented in Figure 1. After screening and filtration, 85

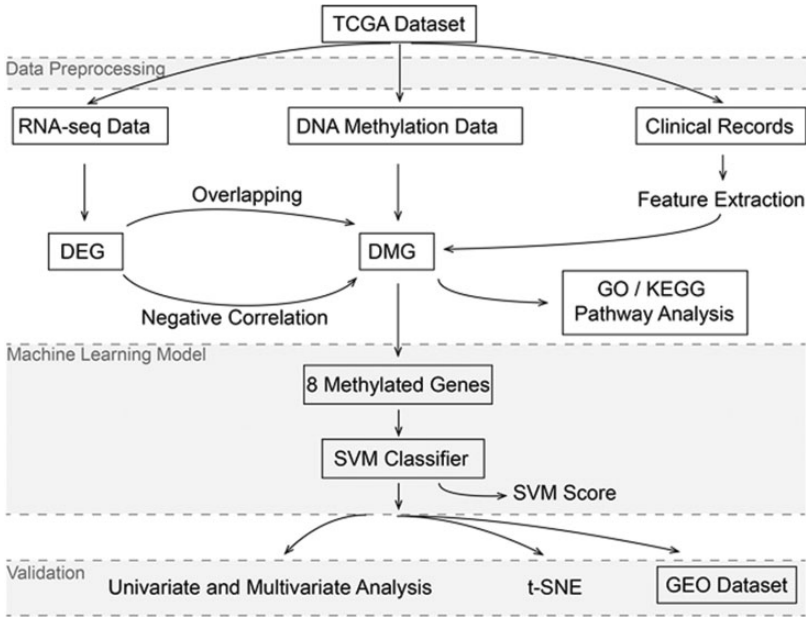


Figure 1. Flowchart of the whole analysis process.

patients who had been followed up for more than 2 years without recurrence were included in the no recurrence group, and 50 patients with evidence of recurrence were included in the recurrence group based on prior findings that recurrence usually occurs in the first 2 years after completing treatment.^{29,30}

Identification of DMGs and DEGs

To identify the DMGs and DEGs between the recurrence and no recurrence groups, we screened 24,278 DNA-methylated genes and 20,104 RNA expression datasets from TCGA database (Figure 2a). Volcano plots were employed to present the distribution of fold changes in DNA methylation or gene expression in each group (Figure 2b, c). In total, 305 DMGs ($P < 0.01$; absolute log fold change > 0.025) and 262 DEGs ($P < 0.01$; absolute log fold change > 0.4) were extracted. Eleven hypermethylated genes and five hypomethylated genes

overlapped with the DEGs, and DNA methylation and mRNA expression levels were negatively correlated ($P < 0.05$) in eight of these genes (*ZFYVE21*, *C17orf62*, *PDIA6*, *CUL7*, *TIGIT*, *MPHOSPH10*, *LOC100132215*, and *CD3E*), as presented in Figure 2d, e. These eight DNA-methylated genes were used to further establish the recurrence prediction model.

The SVM-based prediction model of DNA methylation

The β values of the eight selected DNA methylation sites from 135 patients in TCGA dataset were visualized by a heatmap using the pheatmap package in R (www.r-project.org) (Figure 3a). We trained the SVM model according to the recurrence outcomes (Figure 3b). The confusion matrix is presented in Figure 3c, and the ROC curve (AUC = 0.833) is presented in Figure 3d. The SVM score was generated to predict the recurrence rate with the β values

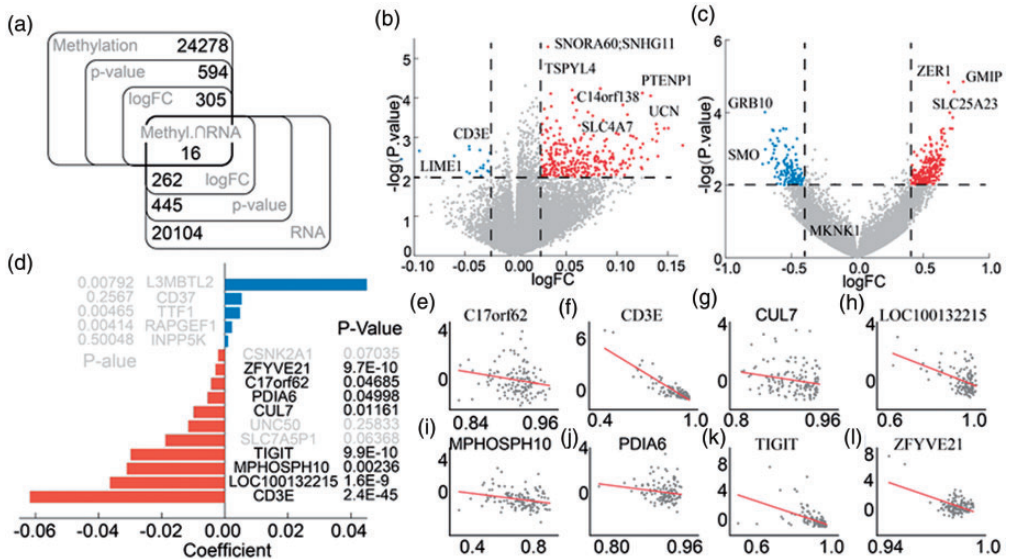


Figure 2. Selection of DMGs in recurrent and nonrecurrent cervical cancer samples. (a) Venn diagram of overlapping DEGs and DMGs. (b) Volcano plot of DMGs. (c) Volcano plot of DEGs. (d) Coefficient distribution of the gene signatures. (e) The correlation between gene expression and DNA methylation levels in cervical cancer.

DMG, differentially methylated gene; DEG, differentially expressed gene.

of the eight DNA-methylated genes using the following formula:

$$\begin{aligned} \text{SVM score} = & 0.48 \times \beta_{C17orf62} - 0.61 \times \beta_{CD3E} \\ & + 0.17 \times \beta_{CUL7} + 0.37 \\ & \times \beta_{LOC100132215} + 0.12 \\ & \times \beta_{MPHOSPH10} + 0.25 \times \beta_{PDIA6} \\ & - 0.49 \times \beta_{TIGIT} + 0.49 \\ & \times \beta_{ZFYVE21} - 0.74 \end{aligned} \quad (2)$$

Univariate and multivariate analyses

The clinical features and SVM scores of these patients were analyzed by univariate analysis (Table 1). The SVM score, FIGO stage, N stage, M stage, and pathologic grade were risk factors for recurrence ($P < 0.05$) in univariate analysis. Multivariate Cox regression analysis was further conducted using these risk factors.

The SVM score (hazard ratio [HR] = 0.42; 95% confidence interval [CI] = 0.26–0.67, $P < 0.001$), and N stage (HR = 3.31; 95% CI = 1.17–9.33, $P < 0.05$) were independent risk factors for recurrence (Table 1).

Visualizing the model using t-SNE

The results of our training model with t-SNE are presented in Figure 4. The eight-dimensional data were mapped onto a two-dimensional space using the dimensionality reduction method. Samples in each group are represented as a data point in red or blue. The results revealed that samples closer to the bottom left were more likely to be from patients in the recurrence group.

External validation of the GEO database

An independent external GEO cohort (GSE30759) was applied to validate the SVM-based recurrence prediction model

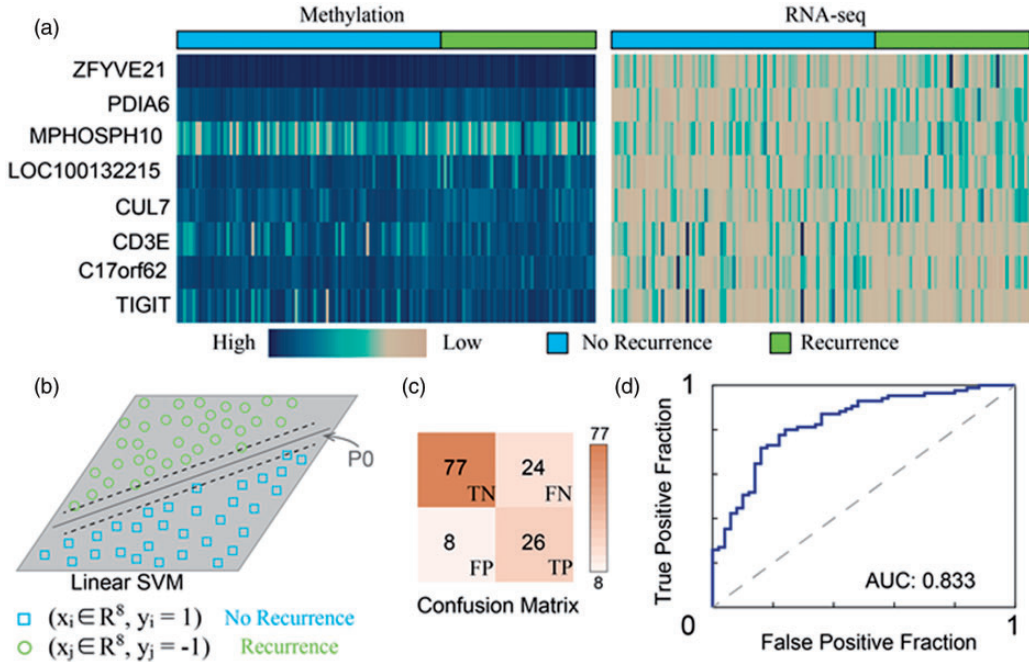


Figure 3. Establishment of the prediction model. (a) Heat map of eight methylation-related genes in patients with cervical cancer. (b) Schematic of the SVM algorithm. (c) The confusion matrix of the prediction model. Positive: recurrence. (d) ROC curve of the prediction model, $AUC = 0.833$. SVM, support vector machine; ROC, receiver operating characteristic; AUC, area under the curve.

of DNA methylation. A flowchart representation of the overall procedure is presented in Figure 5a. Three genes without the probes in the GEO dataset were replaced by the corresponding mean values in TCGA dataset (*LOC100132215*: 0.91; *TIGIT*: 0.91; and *ZFYVE21*: 0.98) to adjust the threshold of the reduced SVM score. As presented in Figure 5b, the methylation levels of the four genes (b1: *PDIA6*; b2: *CUL7*, b3: *CD3E*, and b4: *C17orf62*) with similar distributions in two datasets formed the reduced SVM score. The mean value of *MPHOSPH10* (b5: TCGA: 0.71 ± 0.11 ; GEO: 0.02 ± 0.01) in TCGA dataset was subtracted to further update the reduced SVM score threshold.

As illustrated in Figure 5c, the initial threshold of the SVM score was 0 on the basis of eight genes. The cost of missing

genes was the range of uncertainty around the threshold, in which a sample indicated a high or low risk of recurrence. To improve the identification of high-risk patients, we considered samples within the same region as those from patients at high risk. Therefore, the reduced SVM score formula for the GEO dataset was as follows:

$$\begin{aligned}
 TH_{reduced}^{GEO} = & 0.74 - 0.37 \times \langle \beta_{LOC100132215}^{TCGA,high} \rangle \\
 & + 0.49 \times \langle \beta_{TIGIT}^{TCGA,high} \rangle - 0.49 \\
 & \times \langle \beta_{ZFYVE21}^{TCGA,high} \rangle - 0.12 \\
 & \times \langle \beta_{MPHOSPH10}^{TCGA,high} \rangle
 \end{aligned} \tag{3}$$

In the aforementioned equation, “high” denotes all samples at high risk, and $\langle \dots \rangle$ indicates the average. The reduced SVM score threshold for the four genes in the

Table 1. Stratification analysis for risk factors for recurrence in patients from TCGA

Clinical characteristic	Univariate analysis			Cox regression analysis		
	Mode	No recurrence (N = 85)	Recurrence (N = 50)	HR (95% CI)	P-value	P-value
Age, years [§]	Mean ± SD	48.88 ± 12.37	51.16 ± 13.99		3.31E-01	
SVM score [§]	Mean ± SD	1.32 ± 1.14	-0.21 ± 1.21		2.62E-11	3.02E-04
Clinical characteristic	Group	No recurrence (N = 85)	Recurrence (N = 50)			
FIGO stage [¶]						
	I	48	24		3.09E-02	9.34E-01
	2	23	11			
	3	11	9			
	4	2	5			
	NA	1	1			
Pathologic M stage [¶]						
	M0	39	11		2.15E-02	9.11E-01
	M1	1	4			
	MX	45	35			
Pathologic N stage [¶]						
	N0	46	10		2.50E-03	2.40E-02
	N1	12	13			
	NX	27	27			
Pathologic T stage [¶]						
	T1	45	16		5.55E-02	
	T2	23	9			
	T3	4	2			
	T4	1	4			
	TX	12	19			
Pathologic grade [¶]						
	1	6	3		3.67E-02	
	2	46	16			
	3	26	28			
	NA	6	2			

§: t-test; ¶: Fisher's exact test.

TCGA, The Cancer Genome Atlas; SVM, support vector machine; HR, hazard ratio; CI, confidence interval; SD, standard deviation; NA, not available.

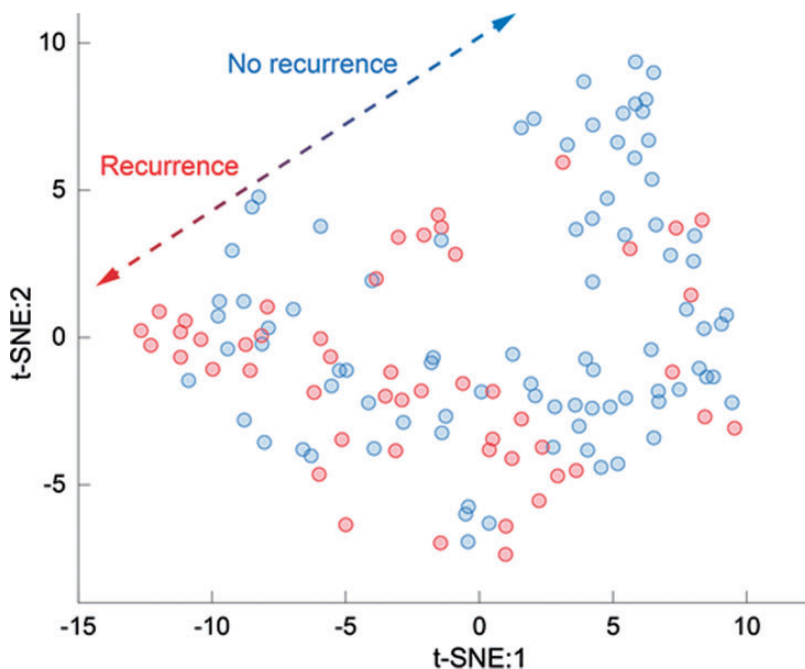


Figure 4. Visualization of the eight-dimensional variables using the t-SNE algorithm. Points in red represent patients with recurrence, and those in blue represent patients without recurrence. t-SNE, t-distributed stochastic neighbor embedding.

GEO dataset $TH_{reduced}^{GEO}$, instead of the initial threshold 0 for the eight genes in the TCGA dataset, equaled 0.304. The β values of the four genes (*CI7orf62*, *CD3E*, *CUL7*, and *PDIA6*) from the GEO dataset then shifted the mean value to 1 in the TCGA dataset, avoiding bias.

Recurrence-free survival was significantly longer in the low-risk group ($TH_{reduced}^{GEO} > 0.304$) than in the high-risk group ($TH_{reduced}^{GEO} \leq 0.304$) in the recurrence curve ($P = 0.039$; Figure 5d). ROC curve analysis was also performed on the prediction result (Figure 5e), and the AUC was 0.773.

GO and KEGG analyses

To research the biological function of DMGs, DAVID was used to perform GO analysis for MFs, BPs, and CCs. MFs

($P < 0.05$; Figure 6a) were enriched in protein binding, poly(A) RNA binding, protein homodimerization activity, and GTPase activator activity. BPs ($P < 0.05$; Figure 6b) were enriched in the regulation of transcription, signal transduction, and positive regulation of GTPase activity immune response. CCs ($P < 0.05$; Figure 6c) were enriched in the plasma membrane, cytosol, and nucleoplasm membrane. The results of KEGG analysis (Figure 6d) suggested that the DMGs were enriched in human T-lymphotropic virus 1 infection, cytokine–cytokine receptor interaction, T-cell receptor signaling (TCR) pathway, and natural killer cell-mediated cytotoxicity.

Discussion

DNA methylation is a major epigenetic mechanism that plays important roles in

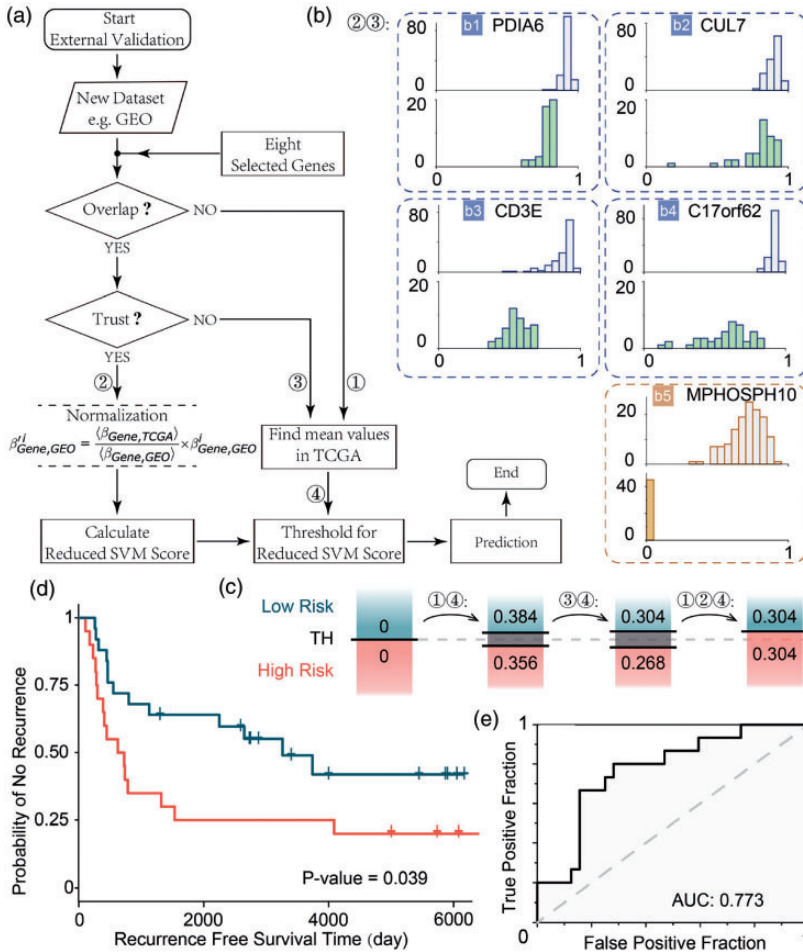


Figure 5. External validation of the GEO dataset. (a) Flowchart for the external validation. (b) The distribution of the methylation levels for the five genes in both TCGA dataset (upper) and the GEO dataset (lower). (c) Threshold shift in the procedure. (d) Survival analysis of the GEO dataset using the prediction model. (e) ROC curve of external validation with the AUC. GEO, Gene Expression Omnibus; TCGA; The Cancer Genome Atlas; ROC, receiver operating characteristic; AUC, area under the curve.

various biological processes, such as the regulation of gene expression,³¹ cell differentiation,³² and inflammation.³³ DNA methylation generally occurs at CpG sites, which are unmethylated or hypomethylated in normal cells.³⁴ Hypermethylation of these CpG sites may silence tumor suppressor genes and lead to carcinogenesis;³⁵ therefore, the identification of abnormally

methylated genes will contribute to disease diagnosis, the prediction of prognosis, and the selection of cancer treatment.

In this study, we comprehensively screened data for patients with cervical cancer from TCGA dataset by comparing patients with and without recurrence. We used the SVM machine-learning algorithm along with the generated SVM scores to

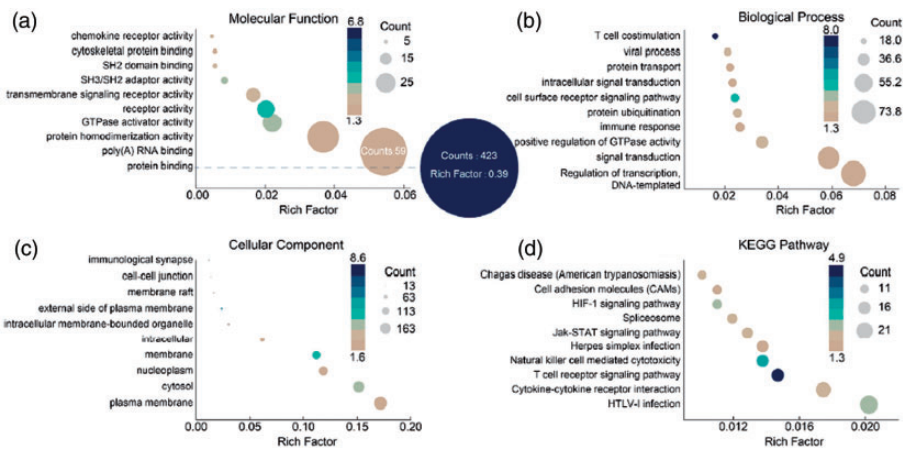


Figure 6. Results of GO and KEGG pathway analyses of DMGs in cervical cancer. The GO terms in (a) MFs, (b) BPs, and (c) CCs. (d) In the results of KEGG pathway analysis, the x-axis represents the rich factor; different colors represent the $-\log_{10}(\text{p-value})$, and the size of the point represents the number of genes. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, molecular function; CC, cellular component.

establish the prediction model. The SVM score was identified as an independent risk factor for recurrence *via* Cox regression analysis. By contrast, FIGO, T, and M stages were not independently predictive of recurrence, which implied that the SVM score was more closely related to tumor recurrence than the traditional tumor staging system. Furthermore, the SVM score is a robust and reliable predictor of recurrence regardless of the pathological type, and it is independent of N staging. The ROC curve illustrated that the model had a high degree of fit. Independent datasets in GEO further supported the model in survival analysis. In clinical applications, this model will allow clinicians to screen patients at high risk of recurrence using eight DNA methylation biomarkers.

The eight DEGs, namely, *ZFYVE21*, *C17orf62*, *PDIA6*, *CUL7*, *CD3E*, *TIGIT*, *MPHOSPH10*, and *LOC100132215*, were found to be differentially affected by DNA methylation between patients with and without recurrence. Some of these genes have been reported to be dysregulated

in cancer or other diseases. *ZFYVE21* is located on chromosome 14 and is associated with the metastasis of colorectal carcinoma³⁶ and malignancy of renal cell carcinoma.³⁷ *C17orf62* regulates NADPH oxidase in phagocytes and contributes to the development of chronic granulomatous disease.³⁸ *CD3E* is an immunoreceptor with a tyrosine-based activation motif (ITAM) in its TCR signal-triggering module that affects overall survival in patients with several cancer types, including cervical cancer.³⁹ *TIGIT* is expressed on the surface of T-cells and natural killer cells and is associated with the reduction in natural kill cell-mediated cytotoxicity and an increase in regulatory T-cell suppression.³⁹ *CUL7* is an oncogene that promotes cancer cell survival by promoting caspase-8 ubiquitination.³⁹ *LOC100132215* is correlated with gene expression in cancerous breast tissues.⁴⁰ *MPHOSPH10* is involved in ribosomal RNA processing during mitosis.⁴¹ *PDIA*⁴⁰ promotes the proliferation and growth of various types of human cancer cells by activating the Wnt/ β -catenin

signaling pathway.³⁹ The role of most of these genes in cervical cancer has not been revealed, and further research is needed to determine their biological functions and mechanisms.

Limitations

There were some limitations in this study. First, to differentiate patients with cervical cancer with and without recurrence, we excluded patients without recurrence who had a short follow-up time, which led to a decrease in the number of samples. Because the follow-up time was not sufficiently long, patients without recurrence within the first 2 years after treatment comprised the no recurrence group, which may cause deviation to some extent. Second, all genes used for modeling methylation from TCGA were not in the external validation datasets from GEO because of the different chips used; therefore, we had to assume that the GEO dataset has the same distribution as the TCGA dataset for *LOC100132215*, *TIGIT*, and *ZFYVE21*. In addition, the histogram (Figure 5b) revealed a large difference in *MPHOSPH10* methylation between the GEO and TCGA datasets. We calculated the average β value in the TCGA dataset and used it in the validation set. Further experimental studies of these genes are needed to clarify their functions and mechanisms in cervical cancer.

Conclusions

In this study, we identified eight recurrence-related DNA methylation biomarkers in cervical cancer through a comprehensive analysis and established a recurrence prediction model using machine learning. This work offers a preliminary but effective approach to predict cancer recurrence using a few biomarkers, which may help clinicians identify high-risk patients and implement active treatment.

Data availability

All analyses and preparation codes, along with the related data files, will be collected and attached in the repository <https://github.com/zfeng2019/8dna>. TCGA Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA-CESC) data were downloaded on January 28, 2016, from the Broad Institute FireBrowse Data Portal (www.firebrowse.org). The GEO dataset is available at <http://www.ncbi.nlm.nih.gov/geo> under the accession number GSE30759.

Declaration of conflicting interest

The authors declare that there is no conflict of interest.


Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions

J.-H. M. and Z. F. designed the study. J.-H. M., Y. H., and L.-Y. L. performed the analysis. J.-H. M. and Z. F. wrote the manuscript. L.-Y. L. and Z. F. visualized the analysis. All authors have read and approved the final manuscript.

ORCID iD

Zhen Feng  <https://orcid.org/0000-0003-1281-4522>

References

1. Bray F, Ferlay J, Soerjomataram I, et al. 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. Peters WA III, Liu PY, Barrett RJ, et al. Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. *Obstet Gynecol Surv* 2000; 55: 491–492.
3. Huang K, Sun H, Li X, et al. Prognostic risk model development and prospective

- validation among patients with cervical cancer stage IB2 to IIB submitted to neoadjuvant chemotherapy. *Sci Rep* 2016; 6: 27568.
4. Obrzut B, Kusy M, Semczuk A, et al. Prediction of 5-year overall survival in cervical cancer patients treated with radical hysterectomy using computational intelligence methods. *BMC Cancer* 2017; 17: 840.
 5. McLachlan J, Boussios S, Okines A, et al. The Impact of Systemic Therapy Beyond First-line Treatment for Advanced Cervical Cancer. *Clin Oncol* 2017; 29: 153–160. doi: <https://doi.org/10.1016/j.clon.2016.10.002>.
 6. Soisson AP, Soper JT, Clarke-Pearson DL, et al. Adjuvant radiotherapy following radical hysterectomy for patients with stage IB and IIA cervical cancer. *Gynecol Oncol* 1990; 37: 390–395.
 7. Sedlis A, Bundy BN, Rotman MZ, et al. A randomized trial of pelvic radiation therapy versus no further therapy in selected patients with stage IB carcinoma of the cervix after radical hysterectomy and pelvic lymphadenectomy: A Gynecologic Oncology Group Study. *Gynecol Oncol* 1999; 73: 177–183.
 8. Koh WJ, Abu-Rustum NR, Bean S, et al. Cervical cancer, version 3.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw* 2019; 17: 64–84.
 9. Moschetta M, Uccello M, Kasenda B, et al. Dynamics of Neutrophils-to-Lymphocyte Ratio Predict Outcomes of PD-1/PD-L1 Blockade. *Biomed Res Int* 2017; 2017: 1506824. doi: 10.1155/2017/1506824.
 10. Belinsky SA, Nikula KJ, Palmisano WA, et al. Aberrant methylation of p16INK4a is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc Natl Acad Sci* 1998; 95: 11891–11896.
 11. Alvarez H, Opalinska J, Zhou L, et al. Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet* 2011; 7: e1001356.
 12. Ali Khan M, Kedhari Sundaram M, Hamza A, et al. Sulforaphane Reverses the Expression of Various Tumor Suppressor Genes by Targeting DNMT3B and HDAC1 in Human Cervical Cancer Cells. *Evid Based Complement Alternat Med* 2015; 2015: 412149. doi: 10.1155/2015/412149.
 13. Bao X, Anastasov N, Wang Y, et al. A novel epigenetic signature for overall survival prediction in patients with breast cancer. *J Transl Med* 2019; 17: 380.
 14. Löf-Öhlin ZM, Sorbe B, Wingren S, et al. Hypermethylation of promoter regions of the APC1A and p16INK4a genes in relation to prognosis and tumor characteristics in cervical cancer patients. *Int J Oncol* 2011; 39: 683–688.
 15. Abudukadeer A, Bakry R, Goebel G, et al. Clinical relevance of CDH1 and CDH13 DNA-methylation in serum of cervical cancer patients. *Int J Mol Sci* 2012; 13: 8353–8363.
 16. Li JY, Huang T, Zhang C, et al. Association between RASSF1A promoter hypermethylation and oncogenic HPV infection status in invasive cervical cancer: a meta-analysis. *Asian Pac J Cancer Prev* 2015; 16: 5749–5754.
 17. Li X and Cai Y. Methylation-based classification of cervical squamous cell carcinoma into two new subclasses differing in immune-related gene expression. *Int J Mol Sci* 2018; 19: 3607.
 18. Li C, Ke J, Liu J, et al. DNA methylation data-based molecular subtype classification related to the prognosis of patients with cervical cancer. *J Cell Biochem* 2019.
 19. Xie F, Dong D, Du N, et al. An 8-gene signature predicts the prognosis of cervical cancer following radiotherapy. *Mol Med Rep* 2019; 20: 2990–3002.
 20. Network CGAR. Integrated genomic and molecular characterization of cervical cancer. *Nature* 2017; 543: 378–384.
 21. Deng M, Brägelmann J, Kryukov I, et al. FirebrowseR: an R client to the Broad Institute's Firehose Pipeline. *Database (Oxford)* 2017; 2017: baw160.
 22. Vapnik V. *The nature of statistical learning theory*. Springer science & business media, 2013.
 23. Meyer D, Dimitriadou E, Hornik K, et al. <https://CRAN.R-project.org/package=e1071>. 2019.
 24. Laszczyńska O, Severo M, Friões F, et al. Validity of the Seattle Heart Failure Model

- for prognosis in a population at low coronary heart disease risk. *J Cardiovasc Med* 2016; 17: 653–658.
25. Van Der Maaten L and Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; 9: 2579–2605.
 26. Krijthe J, Van Der Maaten L and Krijthe MJ. Package ‘Rtsne.’ 2018.
 27. Consortium TGO. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2014; 43: D1049–D1056. doi: 10.1093/nar/gku1179.
 28. Jiao X, Sherman BT, Huang DW, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 2012; 28: 1805–1806.
 29. Pfandler KS and Tewari KS. Changing paradigms in the systemic treatment of advanced cervical cancer. *Am J Obstet Gynecol* 2016; 214: 22–30.
 30. Boussios S, Seraj E, Zarkavelis G, et al. Management of patients with recurrent/advanced cervical cancer beyond first line platinum regimens: Where do we stand? A literature review. *Crit Rev Oncol Hematol* 2016; 108: 164–174.
 31. Thakur C, Chen B, Li L, et al. Loss of mdig expression enhances DNA and histone methylation and metastasis of aggressive breast cancer. *Signal Transduct Target Ther* 2018; 3: 25.
 32. Kulis M, Merkel A, Heath S, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet* 2015; 47: 746.
 33. Ai T, Zhang J, Wang X, et al. DNA methylation profile is associated with the osteogenic potential of three distinct human odontogenic stem cells. *Signal Transduct Target Ther* 2018; 3: 1–8.
 34. Hao X, Luo H, Krawczyk M, et al. DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci* 2017; 114: 7414–7419.
 35. Jones PA and Laird PW. Cancer-epigenetics comes of age. *Nat Genet* 1999; 21: 163–167.
 36. Nagano M, Hoshino D, Koshiba S, et al. ZF21 protein, a regulator of the disassembly of focal adhesions and cancer metastasis, contains a novel noncanonical pleckstrin homology domain. *J Biol Chem* 2011; 286: 31598–31609.
 37. Takahashi M, Tsukamoto Y, Kai T, et al. Downregulation of WDR 20 due to loss of 14q is involved in the malignant transformation of clear cell renal cell carcinoma. *Cancer Sci* 2016; 107: 417–423.
 38. Arnadottir GA, Norddahl GL, Gudmundsdottir S, et al. A homozygous loss-of-function mutation leading to CYBC1 deficiency causes chronic granulomatous disease. *Nat Commun* 2018; 9: 4447.
 39. Wang Q, Li P and Wu W. A systematic analysis of immune genes and overall survival in cancer patients. *BMC Cancer* 2019; 19: 1225.
 40. Moarii M, Boeva V, Vert JP, et al. Changes in correlation between promoter methylation and gene expression in cancer. *BMC Genomics* 2015; 16: 873.
 41. Westendorf JM, Konstantinov KN, Wormsley S, et al. M phase phosphoprotein 10 is a human U3 small nucleolar ribonucleoprotein component. *Mol Biol Cell* 1998; 9: 437–449.