

# A 8-gene-methylation-based Signature for Head and Neck Squamous Cell Carcinoma Prognosis

Zihui Wang

Beijing Chaoyang Hospital <https://orcid.org/0000-0003-4853-1890>

Jingrun Yang

PLAGH: Chinese PLA General Hospital

Lihong Liu (✉ [liulihong@bjcyh.com](mailto:liulihong@bjcyh.com))

Head and neck squamous cell carcinoma, Survival prognosis, Gene expression, DNA methylation

---

## Research

**Keywords:** Head and neck squamous cell carcinoma, Survival prognosis, Gene expression, DNA methylation

**Posted Date:** August 2nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-735108/v1>

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

[Read Full License](#)

---

# Abstract

**Background** Head and neck squamous cell carcinoma (HNSCC), accounting for 6% of systemic malignant tumors, has an increasing incidence rate year by year worldwide. A large number of studies have investigated the tumor markers to determine clinical stage, prognosis, treatment evaluation, predict relapses, and overall survival of HNSCC patients, with controversial results.

**Methods** In this paper, we comprehensively analyzed gene expression and DNA methylation data sets of HNSCC and adjacent non-tumor samples from The Cancer Genome Atlas (TCGA). Univariate cox regression analysis followed by sure independence screening (SIS) were used for identifying differential methylation signatures to stratify patients with significantly different prognosis.

**Results** We identified methylation levels of *HS3ST1*, *TOMM34*, *RPL26L1*, *MTHFD2*, *ORC1*, *MYOSLID*, *UHRF1* and *AL357033.3* as potential HNSCC prognosis signatures, and verified the correlation between their gene expression and the corresponding methylation. Their reliability for predicting the prognosis of HNSCC was confirmed in an independent dataset.

**Conclusions** In conclusion, we built a 8-gene-methylation-based signature which can well assess the prognosis of HNSCC patients.

## 1. Introduction

The neck squamous cell carcinoma (HNSCC) is the most common pathological type of head and neck tumors, which develops in the epithelial layer of the mouth, pharynx and larynx, ranking sixth in systemic malignant tumors (1, 2). Each year, 600,000 HNSCC patients are newly added in the world (3). Due to the hidden anatomical location, difficulty in early detection and strong invasiveness, the overall prognosis of HNSCC is poor (4). Many factors, including smoking, alcohol consumption, viral infection (HPV), are closely associated with increased risk of HNSCC developing (3, 5, 6). However, only a small part of the above population will eventually develop head and neck tumors, suggesting that the critical position of genetic factors (7).

Over the past three decades, despite the tremendous advances in treatments such as surgery, minimally invasive surgery, precision radiotherapy, chemotherapy, and monoclonal antibody therapy, 25–50% of patients will relapse within 3–5 years after the first treatment (8, 9). Thus, early diagnosis is important for improving the survival rate of patients (10). The in-depth studies of biomarkers have provided possibilities for early diagnosis, grading, recurrence and prognosis assessment of HNSCC. Recently, several tumor biomarkers with clinical application have been reported, including matrix metalloproteinases (MMPs) (11, 12), vascular endothelial growth factor (VEGF) (13), interleukin (IL-6 and IL-8) (14).

DNA methylation is the most common epigenetic regulator that affects gene expression and chromatin stability without altering genomic sequences. Carvalho's group found that DNA methylation specificity

was as high as 90% in the serum and saliva samples from amounts of patients with HNSCC (15). Undoubtedly, DNA methylation plays an important role in the occurrence and development of HNSCC. Therefore, we performed comprehensively analyses of gene expression and DNA methylation data to identify key genes affecting the prognosis of HNSCC, and constructed a prognostic prediction model based on these characteristic genes (Fig. 1). Our study should be useful for screening potential therapeutic targets and the development of effective treatment methods of HNSCC.

## **2. Materials And Methods**

### **2.1 Data source**

Datasets on DNA methylation and mRNA expression profiles of HNSCC were obtained from the Cancer Genome Atlas (TCGA, <https://tcgadata.nci.nih.gov/tcga/>). The datasets of TCGA-HNSC Project, which including the information of 527 HNSCC cases, were selected for further analysis. Among them, data of 493 cases with complete information of mRNA expression data (HTSeq-Counts format), DNA methylation data (Platform: Illumina Human Methylation 450 (HM450) arrays) and clinical data was downloaded.

### **2.2 Preprocessing of DNA methylation and mRNA expression data**

The mRNA expression data and DNA methylation level data were obtained from the TCGA database, in which mRNA was raw counts data and DNA methylation levels were expressed as Beta Value. From the 493 patients, 246 cases were randomly selected as training sets to construct a prognostic prediction model based on gene expression levels, and the other 247 cases were used as test sets to verify the reliability of this model.

### **2.3 Gene differentially expressed analysis.**

Forty-three cases with expression data of both tumor and paracancerous normal tissues were selected from the gene expression level matrix of preprocessed samples for screening the differentially expressed genes in HNSCC samples. The differentially expressed genes were screened by using the DESeq2 package in R software using  $|\log_2 \text{Fold Change}| > 4$  and (Benjamini & Hochberg corrected P value)  $BH_p < 0.01$  as a threshold.

### **2.4 Construction of prognostic prediction model.**

Combining the clinical data of 246 patients in the training set, we performed a single factor survival analysis of the differential genes. The significantly differential genes that correlated with prognosis were screened using Kaplan-Meier method in survival R package. Then, the stable and reliable characteristic genes were identified by using the sure independence screening (SIS) with LASSO Cox penalty regression in SIS package. The prognostic prediction model was constructed based on these characteristic genes to assess overall survival of HNSCC samples, and the final predictive model was consisted of the expression levels of the characteristic genes and their multivariate analysis weights.

## **2.5 Verification of prognostic prediction model.**

The scores of each sample in the training set and test set were calculated separately based on the model obtained from the training set. The samples in the training set were divided into two groups based on the scores. The survival curves were performed by the Kaplan-Meier method and comparison between each group were examined by log rank-test. The samples in training set were divided into two groups by selecting an appropriate grouping score threshold, while the samples in test set were divided into two groups with the same score threshold. The difference of overall survival between the two groups were examined by log-rank test. In addition, we performed correlation analysis and single factor survival analysis on the expression data and their corresponding methylation data of the characteristic genes in the model.

## **3. Results**

### **3.1 Clinical data characteristics of the HNSCC sample.**

The clinical data of 493 HNSCC patients were shown in Table 1, including age, gender, ethnicity, and TNM staging. Then a chi-square test was performed for each indicator of the training set and test set, indicating that these two groups were independent of age, gender, race and cancer stage. Thus, the clinical data characteristics had no effect on the construction and verification of the prognostic prediction model.

Table 1  
Clinical characteristics

Characteristic	Training set (n = 246)	Validation set (n = 247)	Chi-square test (P)
<b>Age</b>			0.9593
Range	25–88	20–90	
Median years	62	61	
<b>Gender, n (%)</b>			0.9406
Male	181 (73.6)	180 (72.9)	
Female	65 (26.4)	67 (27.1)	
<b>Race, n (%)</b>			
White	214 (87.0)	208 (84.2)	0.9210
Asian	5 (2.0)	5 (2.0)	
Black or African American	20(8.1)	25 (10.1)	
American Indian or Alaska native	1(0.4)	1(0.4)	
No reported	6 (2.5)	8 (3.3)	
<b>TNM stage, n (%)</b>			0.9868
Early (I-II)	46(18.7)	47 (19.0)	
Advance (III-IV)	167(67.9)	166 (67.2)	
Not reported	33 (13.4)	34 (13.8)	

## 3.2 Differentially expressed genes

The gene expression levels of HNSCC tumor samples and normal samples were analyzed by using the R-based DESeq2 package analyzes. As shown in Fig. 2A, 3682 differentially expressed genes were obtained with a standard of  $|\log_2 \text{Fold Change}| > 4$  and  $BH\_p < 0.01$ . Among these differentially expressed genes, 2928 genes were up-regulated, while 754 genes were down-regulated in HNSCC tumor samples.

## 3.3 SIS model

A total of 983 differentially expressed genes related to overall patient survival were obtained by using the single factor survival analysis based on Kaplan-Meier method, and 8 genes were identified as stable and reliable characteristic genes by using the sure independence screening (SIS) with LASSO Cox penalty regression, including *HS3ST1* (ENSG00000002587.8), *TOMM34* (ENSG00000025772.7), *RPL26L1* (ENSG00000037241.6), *MTHFD2* (ENSG00000065911.10), *ORC1* (ENSG00000085840.11), *MYOSLID*

(ENSG00000229647.1), *UHRF1* (ENSG00000276043.3), and *AL357033.3* (ENSG00000276317.1). The expression of characteristic genes in tumor and normal tissues was shown in Fig. 2B.

The prognostic prediction model based on multiple characteristic genes were constructed which parameters as follows:

Score =  $0.031314301 \times \text{HS3ST1} + 0.011349517 \times \text{TOMM34} + 0.009438558 \times \text{RPL26L1} + 0.012175823 \times \text{MTHFD2} - 0.044957765 \times \text{ORC1} + 0.036879138 \times \text{MYOSLID} - 0.004098338 \times \text{UHRF1} - 0.028640613 \times \text{AL357033.3}$ .

### 3.4 Verification of SIS model

The highest chi-square value of training set obtained from log-rank test for Kaplan-Meier survival was used to divide the samples in training set into high-risk and low-risk groups (HR = 3.34, P value =  $7.14 \times 10^{-11}$ ). As shown in Fig. 3, the higher scores obtained by the model were significantly correlated with the shorter survival rates in the training group. The samples in test set were divided into two groups with the same scoring threshold of training set. Also, the risk of death in the high-risk group was 1.76 times higher than in the low-risk group (P value =  $4.01 \times 10^{-3}$ ). This result indicated that the SIS model we constructed can be used for predicting the prognosis of HNSCC and had high reliability.

Furthermore, A total of 246 samples in the training set had gene expression data and DNA methylation level data, and 4 of the 8 characteristic genes in the SIS model had methylation data of CpG sites in the promoter region, which were *HS3ST1*, *ORC1*, *TOMM34* and *UHRF1*. Thus, we analyzed the correlation between the expression levels of these 4 genes and the methylation levels of their corresponding CpG sites. As shown in Fig. 4, there was no significant correlation between the methylation and gene expression levels of *HS3ST1*, *ORC1* and *TOMM34*, however, the methylation level of *UHRF1* was significantly positively correlated with gene expression level. In addition, we further analyzed the relationship between the expression of these 4 genes and the overall survival of the patients, and the survival curves showed significant differences between high-risk and low-risk groups (HR = 1.67, P value =  $5.58 \times 10^{-3}$ ) according to the differently expression of *TOMM34*.

In this experiment, the correlation between the expression and corresponding methylation of characteristic genes was not consistent with the expected, but in general, the prognostic model we constructed had certain value for the prognosis of HNSCC.

## 4. Discussion

The development of cancer often involved complicated regulatory networks, and many studies have shown that DNA methylation in malignant tumor cells is abnormal. Chang et al found that the degree of methylation of p15 and p16 in body fluids and tumor tissues of HNSCC patients were significantly higher than these of healthy group (16). The methylation level of p15 in HNSCC tissues of patients with long-term smoking or drinking was higher than non-smoking or non-drinking patients, indicating that the

abnormal DNA methylation is significantly correlated with clinical high-risk factors of HNSCC (17). Besides, Sanchez-Cespedes's group have detected the methylation level of p16,6-oxomethylguanine-DNA methyltransferase, a death-related protein kinase in HNSCC tissues, suggesting that DNA methylation existed in all pathological stages of tumor tissues (Sanchez-Cespedes, 2000). In this study, we developed a prognostic model for HNSCC that analyzed the transcriptomes and DNA methylation data of 493 HNSCC samples. The results showed that prognostic signature was significantly associated with overall survival in HNSCC patients, and patients with higher model scores tended to have poorer survival. 8 genes were identified as stable and reliable characteristic genes in SIS model, including *HS3ST1*, *TOMM34*, *RPL26L1*, *MTHFD2*, *ORC1*, *MYOSLID*, *UHRF1*, and *AL357033.3*.

It was found that the methylation level of UHRF1 showed a significant related with the gene expression level, and the expression level of the gene TOMM34 was significant different between high-risk and low-risk groups. In the present study, it had been reported that Ubiquitin-like PHD and Ring Finger domain 1 (UHRF1) genes were newly discovered nuclear protein genes closely related to cell growth, and it played an important role in regulating biological processes such as DNA damage repair, cell proliferation, cell cycle, and apoptosis (18, 19). Furthermore, UHRF1 is also an important epigenetic regulator maintaining DNA methylation and histone code in the cell, involving in the regulation of tumorigenesis and progression (20, 21). Several studies have suggested UHRF1 as a potential universal biomarker for cancers (22–25). Translocase of the outer mitochondrial membrane 34 (TOMM34) gene transcript as one of the tops differentially expressed gene, have been proven to be associated with features of aggressive behavior including higher tumor grade, advanced nodal stage, larger tumor size and lymphovascular invasion (26). The prognostic value of TOMM34 has been reported in colorectal cancer (27, 28).

Moreover, folate metabolism was central to cell proliferation and a target of commonly used cancer chemotherapeutics. In particular, the mitochondrial folate-coupled metabolism was thought to be important for proliferating cancer cells (29). The mitochondrial enzymes bifunctional methylenetetrahydrofolate dehydrogenase / cyclohydrolase (MTHFD2) in this pathway was highly expressed in human tumors, and broadly required in survival of cancer cells. Philip M et al revealed that overexpression of MTHFD2 was associated with both high proliferation rates and c-MYC overexpression (30). It has been reported that the overexpression of MTHFD2 was associated with poor prognosis of breast cancer patients and with an increased rate of invasion and metastasis (31). As MTHFD2 is over expressed in rapidly replicating tumor cells but not in adult tissue, it is suitable as a therapeutic target for selective cancer treatment (32).

In summary, the above reports on the functions of our characteristic genes were consistent with our finding, however, some genes were shown to be related with the cardiovascular disease, such as *HS3ST1* and *MYOSLID*, and other genes had been less studied. Thus, we would verify the relationship between these genes and HNSCC as well as the role of genes in the signaling pathway in the future study. In addition, we should expand the clinical data in further investigate to better assess the survival of HNSCC and obtain the prognostic model.

## **5. Conclusion**

In conclusion, we obtained an 8-genes-based prognostic model of HNSCC by comprehensive bioinformatics analysis of DNA methylation and gene expression data, and verified its prognostic value of patients with HNSCC. This study should be helpful for clinical treatment and experimental research of HNSCC.

## **Declarations**

## **Funding:**

No funding was received for the creation of this article.

## **Acknowledgments:**

The authors are grateful to Xinhua Liu for proofreading this manuscript.

## **Data Availability Statement:**

All data generated or analyzed during this study are included in this published article.

## **Authors' contributions:**

Zihui Wang made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data; Jingrun Yang performed the experiments and involved in drafting the manuscript and revising it critically for important intellectual content. Lihong Liu agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## **Conflict of Interest:**

The authors declare that they have no competing interests, and all authors should confirm its accuracy.

## **Ethics approval and consent to participate:**

Not applicable.

## **Consent for publication:**

Not applicable.

## Data citation

The data that support the findings of this study are available in [TCGA-HNSC] at [<https://tcgadata.nci.nih.gov/tcga/>]. These data were derived from the following resources available in the public domain: [TCGA-HNSC, <https://tcgadata.nci.nih.gov/tcga/>] Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012.

## References

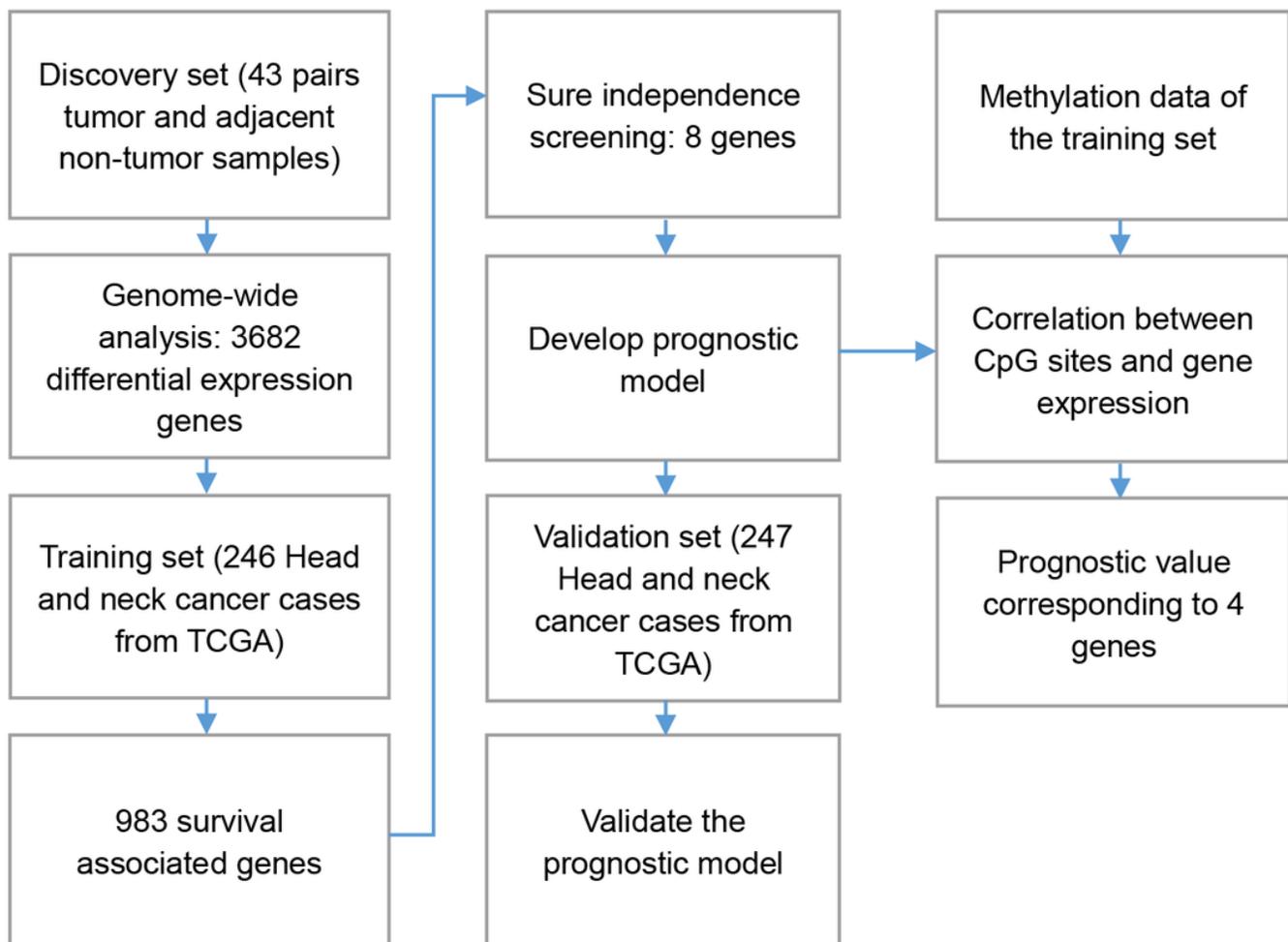
1. Schmitz S, Machiels JP. Molecular biology of squamous cell carcinoma of the head and neck: relevance and therapeutic implications. *Expert Rev Anticancer Ther.* 2010;10:1471–84.
2. Zhao M, Sano D, Pickering CR. Assembly and initial characterization of a panel of 85 genomically validated cell lines from diverse head and neck tumor sites. *Clin Cancer Res.* 2011;17:7248–64.
3. Rezende TMB, Freire MDS, Octávio. Luiz Franco: Head and neck cancer: proteomic advances and biomarker achievements. *Cancer.* 2010;116:4914–25.
4. Siegel R, Ward E, Brawley O. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *Ca Cancer J Clin.* 2011;61:212–36.
5. Thompson, Trisha L. Factors associated with mortality in 2-year survivors of head and neck cancer. *Archives of Otolaryngology–Head Neck Surgery.* 2011;137:1100.
6. Lewin F, Norel SE, Johansson H. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck: a population-based case-referent study in sweden. *Cancer.* 1998;82:1367–75.
7. Gillison ML, D'Souza G, Westra W. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *JNCI Journal of the National Cancer Institute.* 2008;100:407–20.
8. Bernier J. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl J Med.* 2004;350:1945–52.
9. Cooper JS, Pajak TF, Forastiere AA. Postoperative Concurrent Radiotherapy and Chemotherapy for High-Risk Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med.* 2004;350:1937–44.
10. Mydlarz WK, Hennessey PT, Califano JA. Advances and perspectives in the molecular diagnosis of head and neck cancer. *Expert Opinion on Medical Diagnostics.* 2010;4:53–65.
11. Hong SD, Hong SP, Lee JI. Expression of matrix metalloproteinase-2 and – 9 in oral squamous cell carcinomas with regard to the metastatic potential. *Oral Oncol.* 2000;36:207–13.
12. Kuropkat C, Duenne AA, Herz U. Significant correlation of matrix metalloproteinase and macrophage colony-stimulating factor serum concentrations in patients with head and neck cancer. *Neoplasma.* 2004;51:375–8.

13. Kyzas PA. Prognostic significance of vascular endothelial growth factor immunohistochemical expression in head and neck squamous cell carcinoma: a meta-analysis. *Clin Cancer Res.* 2005;11:1434–40.
14. Mineta H, Miura K, Ogino T. Prognostic value of vascular endothelial growth factor (vegf) in head and neck squamous cell carcinomas. *Br J Cancer.* 2000;83:775–81.
15. Carvalho AL, Jeronimo C, Kim MM. Evaluation of promoter hypermethylation detection in body fluids as a screening/diagnosis tool for head and neck squamous cell carcinoma. *Clin Cancer Res.* 2008;14:97–107.
16. Chang HW, Ling GS, Wei WI. Smoking and drinking can induce p15 methylation in the upper aerodigestive tract of healthy individuals and patients with head and neck squamous cell carcinoma. *Cancer.* 2010;101:125–32.
17. Kim DH, Nelson HH, Wiencke JK. P16ink4a and histology-specific methylation of cpg islands by exposure to tobacco smoke in non-small cell lung cancer. *Can Res.* 2001;61:3419–24.
18. Zhang ZY, Cai JJ, Hong J. Clinicopathological analysis of uhrf1 expression in medulloblastoma tissues and its regulation on tumor cell proliferation. *Med Oncol.* 2016;33:99.
19. Qin Y, Wang J, Gong W. Uhrf1 depletion suppresses growth of gallbladder cancer cells through induction of apoptosis and cell cycle arrest. *Oncol Rep.* 2014;31:2635–43.
20. Nakamura K, Baba Y, Kosumi K. Uhrf1 regulates global dna hypomethylation and is associated with poor prognosis in esophageal squamous cell carcinoma. *Oncotarget.* 2016;7:57821–31.
21. Myrianthopoulos V, Cartron PF, Matulis D. Tandem virtual screening targeting the sra domain of uhrf1 identifies a novel chemical tool modulating dna methylation. *Eur J Med Chem.* 2016;114:390–6.
22. Saidi S, Popov Z, Janevska V. Overexpression of uhrf1 gene correlates with the major clinicopathological parameters in urinary bladder cancer. *International braz j urol.* 2017;43:224–9.
23. Abu-Alainin W, Gana T, Liloglou T. Uhrf1 regulation of the keap1-nrf2 pathway in pancreatic cancer contributes to oncogenesis. *J Pathol.* 2016;238:423–33.
24. Wan X, Yang S, Huang W. Uhrf1 overexpression is involved in cell proliferation and biochemical recurrence in prostate cancer after radical prostatectomy. *Journal of Experimental Clinical Cancer Research.* 2016;35:34.
25. Liu X, Ou H, Xiang L. Elevated uhrf1 expression contributes to poor prognosis by promoting cell proliferation and metastasis in hepatocellular carcinoma. *Oncotarget.* 2017;8:10510–22.
26. Aleskandarany MA, Negm OH, Rakha EA. Tom34 expression in early invasive breast cancer: a biomarker associated with poor outcome. *Breast Cancer Res Treat.* 2012;136:419–27.
27. Shimokawa T, Matsushima S, Tsunoda T. Identification of tom34, which shows elevated expression in the majority of human colon cancers, as a novel drug target. *Int J Oncol.* 2006;29:381–6.
28. Matsushita N, Yamamoto S, Inoue Y. Rt-qpcr analysis of the tumor antigens tom34 and rnf43 in samples extracted from paraffin-embedded specimens of colorectal cancer. *Oncology Letters.*

2017;14:2281–7.

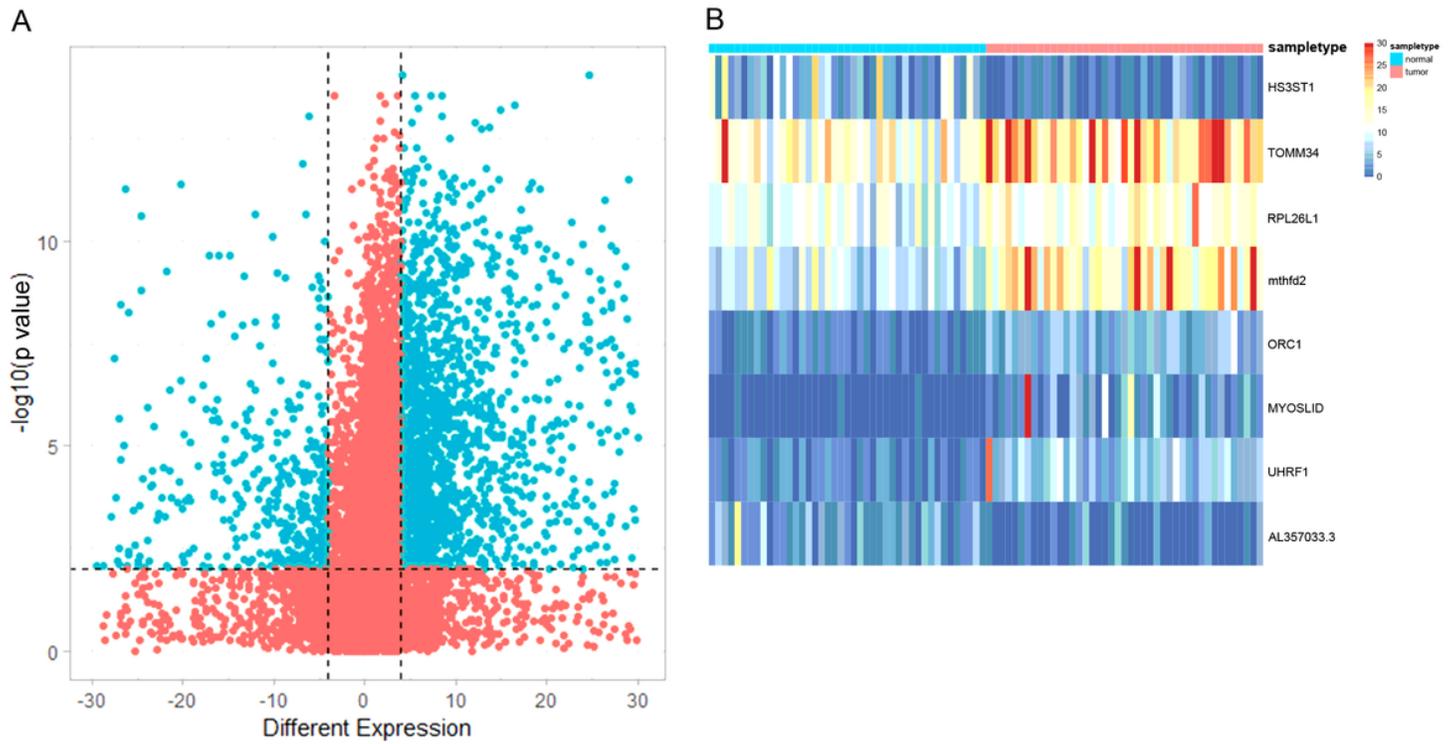
29. Gustafsson Sheppard N, Jarl L, Mahadessian D. The folate-coupled enzyme mthfd2 is a nuclear protein and promotes cell proliferation. *Sci Rep*. 2015;5:15029.
30. Tedeschi PM, Scotto KW, Kerrigan J. Mthfd2- a new twist? *Oncotarget*. 2016;7:7368.
31. Lincet H, Icard P. How do glycolytic enzymes favour cancer cell proliferation by nonmetabolic functions? *Oncogene*. 2015;34:3751–9.
32. Nilsson R, Jain M, Madhusudhan N. Metabolic enzyme expression highlights a key role for mthfd2 and the mitochondrial folate pathway in cancer. *Nat Commun*. 2014;5:3128.

## Figures



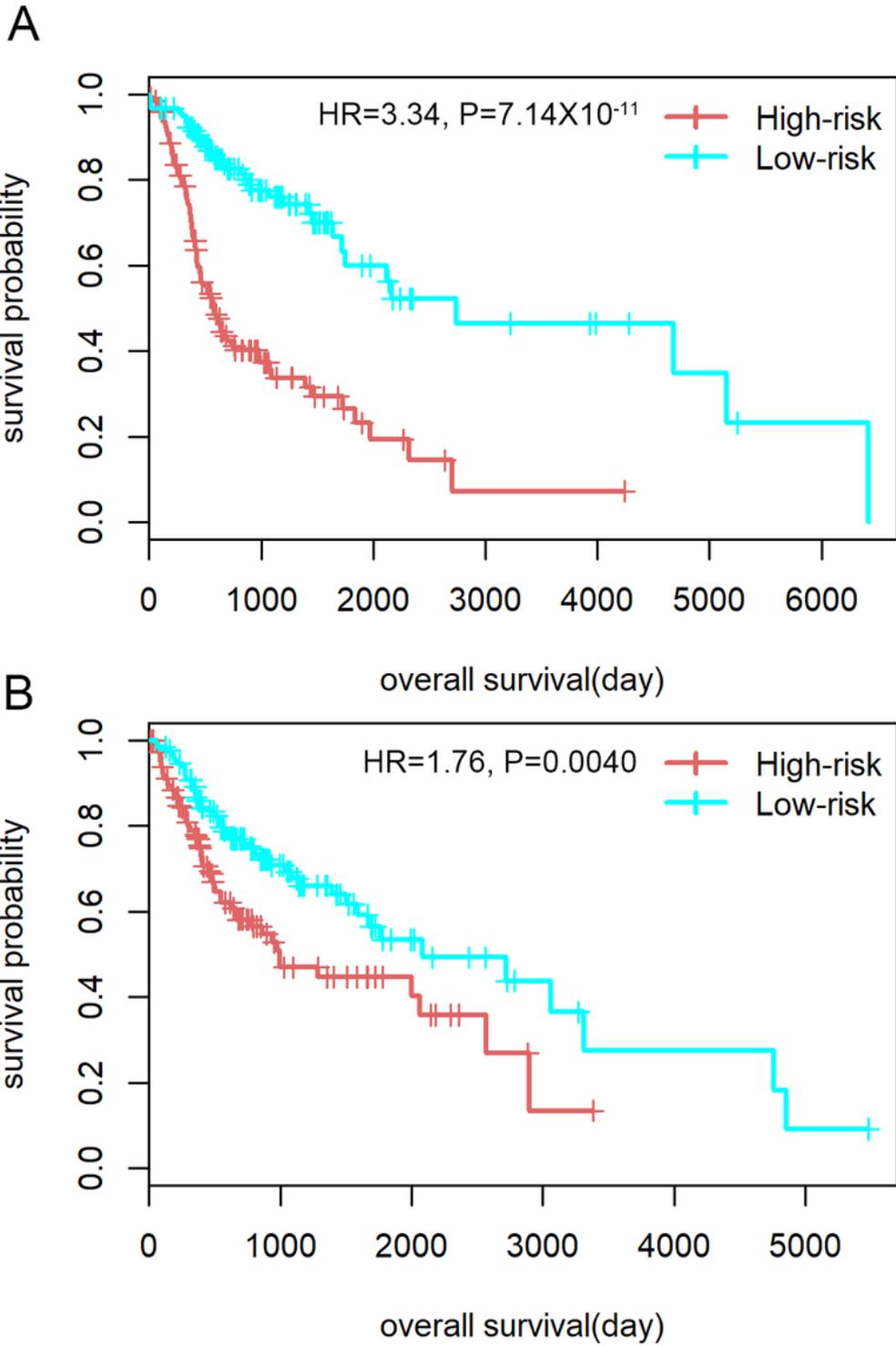
**Figure 1**

A flow chart showing the analysis process of this study.



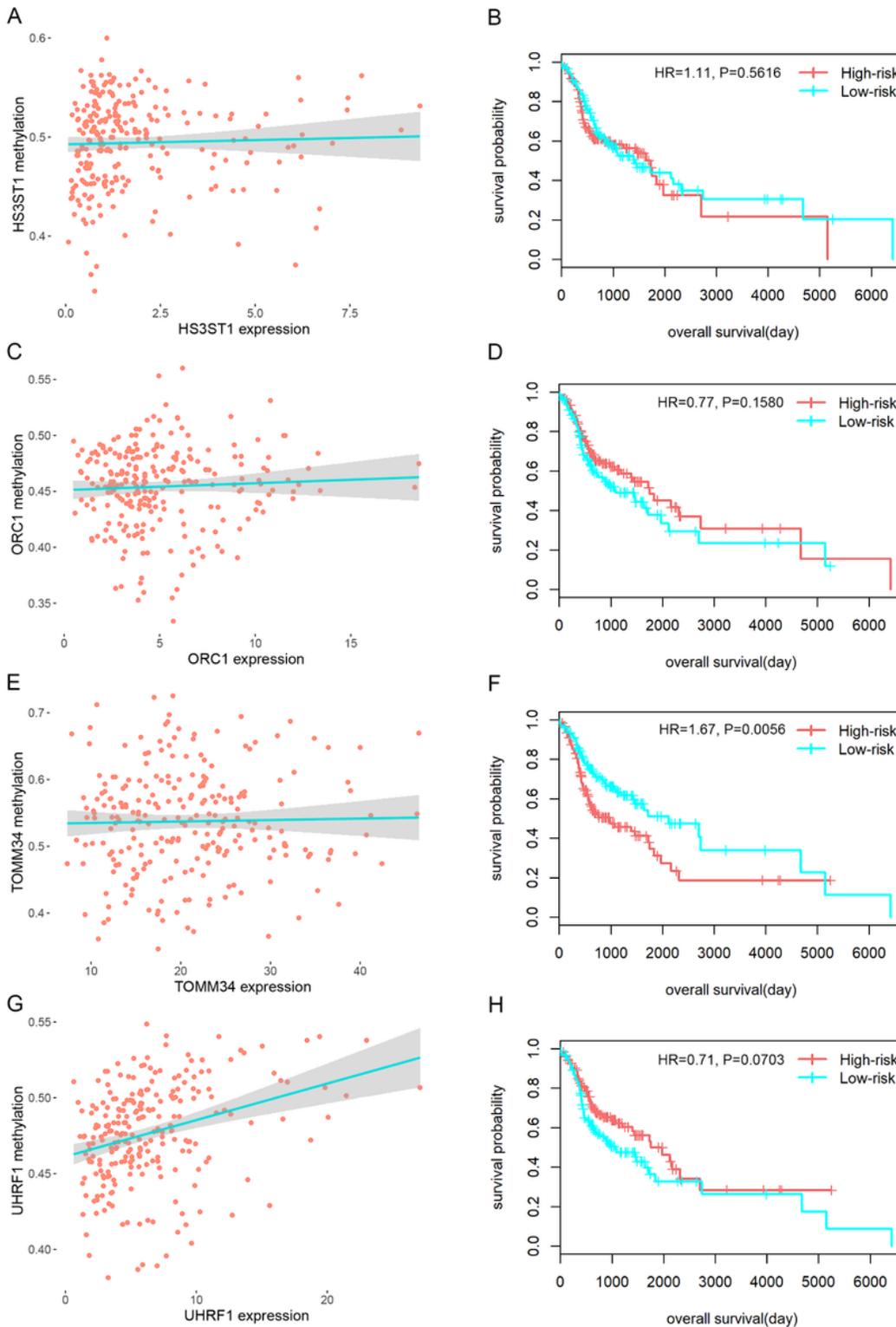
**Figure 2**

Identification of differentially expressed genes in HNSCC. A. The differentially expressed gene results. Blue dots indicated differentially expressed genes, and red dot indicated genes did not change significantly. B. The expressions of characteristic genes in SIS model in tumor and normal tissues.



**Figure 3**

Verification of SIS model. A. Survival curve of training set (HR = 3.34, P value =  $7.14 \times 10^{-11}$ ), B. Survival curve of test set (HR = 1.76, P value =  $4.01 \times 10^{-3}$ ).



**Figure 4**

Verification of characteristic genes in SIS model. A. Correlation between the expression of HS3ST1 and methylation of corresponding CpG sites. B. The relationship of survival curve of HNSCC patients to the expression of HS3ST1. C. Correlation between the expression of ORC1 and methylation of corresponding CpG sites. D. The relationship of survival curve of HNSCC patients to the expression of ORC1. E. Correlation between the expression of TOMM34 and methylation of corresponding CpG sites. F. The

relationship of survival curve of HNSCC patients to the expression of TOMM34. G. Correlation between the expression of UHRF1 and methylation of corresponding CpG sites. H. The relationship of survival curve of HNSCC patients to the expression of UHRF1.