Molecular Subtypes Based on DNA Methylation Profiles can Predict the Prognosis of Patients with Laryngeal Cancer

YingPo Yang
Zunyi Medical University
Qiang Li
Zunyi Medical University
Jukun Song
Guizhou Provincial People's Hospital https://orcid.org/0000-0003-2542-9340
Wei An
Zunyi Medical University
Zhaohui Liu (rzent@163.com)
Zunyi Medical University

Research

Keywords: laryngeal cancer, unsupervised clustering, molecular subtype, prognosis

Posted Date: October 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-95395/v1

Abstract

Background: Laryngeal cancer is a common malignant tumor, with the highest mortality rate and lymph node metastasis rate among cancers. Although laryngeal cancer subtypes have been defined for clinical relevance, there are few studies on the molecular characteristics of laryngeal cancer subtypes.

Methods: DNA methylation profiles of laryngeal cancer were downloaded from The Cancer Genome Atlas database. The univariate Cox regression analysis and multivariate Cox regression analysis was employed to screen the prognosis-related DNA methylation site. Meanwhile, we performed unsupervised clustering using prognostic DNA methylation data to identify the molecular characteristics of laryngeal cancer. We also explored the correlation between Differences in DNA methylation levels and differences in T, N, and M category, age, stage, and prognosis. Functional enrichment analysis was conducted in each molecular subtype. A risk model was constructed via prognostic DNA methylation levels.

Results: DNA methylation data in 123 laryngeal cancer samples were obtained. Through univariate Cox regression analysis, the 17,751 DNA methylation sites were defined as prognostic CpG sites with the threshold of P value less than 0.05. Then, the multivariate Cox regression analysis was further screen the prognosis-related DNA methylation sites. Five subgroups were identified based on consensus clustering using 455 prognostic CpG sites that were significantly associated with patient's survival. We built a prediction model and used it to classify the samples.

Conclusion: The unsupervised clustering was performed using prognostic DNA methylation data and described the seven laryngeal cancer molecular subtypes. Our results provide novel insights into mechanisms and more effective personalized treatments of laryngeal cancer.

Introduction

Laryngeal carcinoma (LC) is one of the most common types in the head and neck region[1]. Laryngeal cancer is expected to account for 17,950 new cases and 3,640 deaths in the United States in 2020[2]. About 0.8% of all new cancer cases and 0.6% of all cancer deaths occur in patients with laryngeal cancer. With the decrease in tobacco use, the incidence of laryngeal cancer has decreased by 2.4% every year in the past 10 years. Although surgery, radiotherapy, and chemotherapy have improved, the prognosis of most patients with laryngeal cancer has not improved significantly in the past thirty years, and the 5-year survival rate is still about 60%[3–5]. It is reported that the recurrence rate of advanced laryngeal cancer is between 25%-50%[6]. Laryngeal cancer is highly heterogeneous, and differences in sensitivity to chemotherapy between clinical subtypes can lead to multiple prognoses. Therefore, it is important to find molecular subtypes for patients with laryngeal cancer, which will help improve the prognosis and treatment effect.

Epigenetics is considered to be heritable changes in gene expression, not related to changes in the DNA sequence. It plays a vital role in carcinogenesis[7–10]. At present, with the development of high-throughput gene detection technologies (especially gene chips and RNA sequencing, DNA methylation), DNA methylation of several promoter sequences, including EZH2, SSTR2, CDKN2A, MGMT, MLH1, and DAPK, has been associated with the onset, development, and malignant transformation of laryngeal cancer[11–13]. The identification of specific epigenetic markers from samples of patients with laryngeal cancer may help to develop personalized treatment plans. Such biomarkers can play a key role in prognostic evaluation, staging, recurrence prediction, and timely initiation of appropriate therapeutic drugs and interventions.

The traditional morphology-based classification system of laryngeal cancer include the occurrence site (primary and secondary laryngeal cancer), histopathological classification (squamous cell carcinomas (SCC), non-SCC cancers),

HPV infection status (human papillomavirus-initiated (HPV+) and HPV-negative (HPV −) disease)[14–16]. Meanwhile, the TCGA analysis identified laryngeal cancer subclasses as atypical, classic, basal, and mesenchymal types based on mRNA profiling. Additionally, other studies based on mutation data has been conducted to identify NSD1- and NSD2-subtypes of laryngeal cancer [15]. Currently, there is a lack of reliable molecular prognostic biomarkers for laryngeal cancer stratification based on DNA methylation profiles. Therefore, the prognostic prediction model based on high-throughput omics data integrates multiple DNA methylation biomarkers to improve clinical prognosis evaluation and personalized treatment in research. This classification system also can help identify molecular subtypes or new laryngeal cancer markers to more accurately classify patients with laryngeal cancer.

Materials And Methods

Data source and data pre-processing

RNA-sequencing data from 129 laryngeal cancer samples were downloaded from the TCGA data portal (https://cancergenome.nih.gov/)[17]. The clinical information of these samples is shown in Supplementary Fig. 1, including follow-up data of 117 patients. The methylation data of the HumanMethylation450 BeadChip array on samples from 129 patients were downloaded from the UCSC Cancer Browser.

The methylation level of each probe is represented by a β value from 0 to 1, corresponding to unmethylated and fully methylated, respectively. Probes with missing data in more than 70% of the samples were removed. The remaining probes that were not available (NAs) were imputed using the k-nearest neighbors (KNN) imputation procedure. The ComBat algorithm in the “sva” of R package was used to eliminate batch effects by combining patient ID information and batches and integrating all DNA methylation array data[18]. Finally, we selected samples for which gene expression profiles were available. In total, 112 samples and 20,6635 methylation sites were included in subsequent analysis.

Identification of molecular subtype characteristics using Univariate COX regression analysis

The univariate Cox regression analysis was used to screen the prognostic CpG sites with a cutoff point of P value less than 0.05. Then, multivariate COX proportional risk regression model was further employed to identify the prognostic related CpG sites, including tumor stage and age as covariates. Finally, the CpG sites that were still significant were used as classification features. The Coxph function in the survival software package R was used to fit the Cox proportional hazard model to the methylation level of CpG, and clinical and demographic features (T category, N category, M category, age, and stage) were used as covariates in the multivariate analysis with the threshold of P value less than 0.001.

Molecular subtype classification using consensus clustering

Consensus clustering was performed with the ConsensusClusterPlus package in R to determine subgroups of laryngeal cancer based on prognostic CpG sites[19]. After executing unsupervised consensus clustering, the item-consensus results and cluster consensus were obtained. Cluster counts of 2, 3, 4, 5, 6, and 7 are evaluated. The cumulative distribution function (CDF), the proportion of ambiguously clustered pairs (PAC), principal component analysis (PCA), and consensus heatmaps were used to assess the optimal K. The category number was selected as the area under the CDF curve and showed no significant change. The heat map corresponding to the consensus clustering was generated by pheatmap R package.

Survival and Clinical Characteristic Analysis

The Kaplan–Meier plots were used to illustrate the overall survival rate among molecular subgroups defined by the DNA methylation profiles. The log-rank test was used to test the significance of differences among the clusters. The chi-square test was used to analyze the association between clinical and biological characteristics and DNA methylation clustering. All tests are two-sided. A P value < 0.05 was considered statistically significant.

Functional enrichment analysis

To identify differentially expressed genes (DEGs) in subtypes, differential expression analysis of subtypes was performed. Gene Ontology (GO) biological process terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were performed with the R package “clusterprofile”[20]. Biological process (BP), cellular component (CC), and molecular function (MF) are the three parts of GO analysis. P-value < 0.05 and false discovery rate (FDR) of < 0.05 were the threshold for GO and KEGG enrichment analysis.

Construction and testing of the prognostic prediction model

With the candidate methylation sites identified by Cox Proportional Hazard Model, we built a prognostic prediction model. The regression coefficients from the multi Cox regression analysis were used to create a classification index for each sample, and used the following formula to weight the expression value of the selected methylation sites: The Prognostic Index = (methylation level of cg1 * Coef1) + (methylation level of cg12* Coef2) + (methylation level of cg3* Coef3) +…+ (methylation level of cgn * Coefn). The KM survival curve, ROC curve, and risk curve were plotted. Meanwhile, the univariate and multivariate COX prognostic analysis was used to determine whether the established prognostic prediction model was an independent prognostic factor for HNSC.

Results

Identification of prognostic related methylation sites

After patient data were preprocessed as described above, 206,635 methylation sites were identified. To identify the specific CpG sites that were significantly correlated with overall survival in laryngeal cancer, the univariate COX regression analysis was conducted. Among these methylation sites, 17,751 methylation sites were identified as potential DNA methylation biomarkers for overall survival in patients with laryngeal cancer. Subsequently, a multivariate COX regression analysis was performed on these CpG sites (variables also include TMN stage, age, stage), and 455 independent CpG sites related to the prognosis of patients with laryngeal cancer were identified (Supplementary Table 1).

Molecular subtype classification based on Consensus Clustering

Using the ConsensusClusterplus R software package, the methylation profiles of 455 CpG sites from 112 samples were used for consensus clustering of samples to obtain molecular subtypes of laryngeal cancer. Consensus unsupervised clustering of 122 samples from laryngeal cancer patients revealed 2–7 clusters. Compared with 3, 4, and 6 clusters, cluster 5 had a lower value for the proportion of ambiguously clustered pairs (PAC), which reflected a near-perfect stable partitioning of the samples at the correct K value (Fig. 1A-1C). The consensus matrix shown in Fig. 2A represents the consensus for k = 5 and displays a well-defined 5-block structure. A heatmap corresponding to the dendrogram in Fig. 2A with T category, N category, M category, stage, age, and DNA methylation subgroup as the annotations is shown in Fig. 2B.

Kaplan-Meier survival analysis showed a significant difference in the prognosis between the five subgroups (P < 0.05). As shown in Fig. 3A, cluster 3 has the best prognosis, while cluster 4 and 2 have the worst prognosis. Then, we analyzed the intra-cluster ratios of 5 clusters according to the T, N, M, stage, and age, as shown in Fig. 3B-3F. The

correlation trends between features and specific clusters are as follows: Cluster 3,4 with young patients; Cluster 3 with the male population; Clusters 2 with higher grade and advance stage; Cluster 3,5 with early-stage; Cluster 4 with higher M stage; Cluster 1,3 with higher T stage; Cluster 3 with lower N stage (Fig. 4F). These results indicate that each clinical parameter is related to a different intra-cluster ratio.

Annotate prognostic-related CpG sites and functional enrichment analysis

Genomic annotation of the above-mentioned prognostic CpG sites was used to identify the promoter regions of 567 genes. Then we conducted a functional enrichment analysis of 567 genes and identified significantly 405 GO terms and 17 enriched pathways ($P < 0.05$), as shown in Fig. 3A, B. The top 5 enriched GO terms were enriched in signal transduction by p53 class mediator, autophagy, DNA repair complex and p53 binding, which included basic cancer-related biological processes. The three most significantly enriched pathways were the TGF – beta signaling pathway, autophagy, and cell cycle. As shown in Fig. 3C, close relationships were identified among the 17 pathways and between those pathways and the cell cycle and cellular senescence.

Then, the expression of methylated genes identified in the subgroups was explored. The expression values of 567 genes out of 122 samples are available. The gene expression heat map is shown in Fig. 3D. The gene expression patterns differ between these subgroups, which indicates that the DNA methylation level usually reflects the expression of these genes.

Identifying subtype-specific CpG sites

The expression of methylated genes identified in the subgroup was explored. 455 CpG sites were analyzed for differences among 5 molecular subtypes, and 70 molecular subtype-specific CpG sites were identified. Among them, subtype 3 has the most differentially expressed CpG sites, which are hypermethylated compared with other subtypes (Table 1). The gene expression heatmap is shown in Fig. 4C. The differences in gene expression patterns between subgroups suggest that DNA methylation levels usually reflect the expression of these genes. Then, we further screen for cluster-specific methylation sites by including the methylation sites as features of clusters. Finally, the 32 cluster-specific methylation sites were identified and the heat map was shown in Fig. 5A. Analysis using clusterProfiler showed that these genes were enriched in 3 pathways, as shown in Fig. 5B. Cluster 3 had the largest number of specific sites, all of which were hypermethylated, and the methylation level was the highest among all the clusters (Fig. 6).

Table 1
Multivariate COX stepwise regression analysis of hub DNA methylation sites.

DNA methylation site	coef	HR	HR.95L	HR.95H	pvalue
cg00413617	-3.90264	0.020189	0.000579	0.704155	0.031278
cg01330280	8.118517	3356.042	19.0643	590790.9	0.002089
cg05893785	5.122021	167.6739	9.392133	2993.413	0.000496
cg07098752	-4.02448	0.017873	0.000666	0.479747	0.016506
cg10699171	3.121739	22.68581	2.139974	240.4916	0.009555
cg16508480	2.441683	11.49237	0.722159	182.8884	0.083736
cg18470295	4.456162	86.15617	2.747028	2702.151	0.011252
cg24464397	-3.47284	0.031029	0.000574	1.678696	0.08809
cg24911113	-3.74838	0.023556	0.001724	0.321816	0.004956

Establishment and evaluation of a prognostic prediction model for laryngeal cancer patients

Cluster 3 was chosen as the candidate cluster because it contains a large number of samples, has a good prognosis, and has the most specific methylation sites. In multivariate Cox regression method, nine methylation sites as optimal features were ultimately recognized, including cg00413617, cg01330280, cg05893785, cg07098752, cg10699171, cg16508480, cg18470295, cg24464397 and cg24911113 (Table 2). Then, we obtained the risk score based on the formula described above and calculated the risk score of each sample (Fig. 7A). As the risk score increased, the methylation level of nine specific sites increased significantly. Hypermethylation was associated with low-risk patients, while hypomethylation was associated with high-risk patients. The area under the curve of the ROC curve for 1,3,5 year reached 0.80, 0.86, and 0.89, respectively, indicating that the model functioned well (Fig. 7B). Based on the median of risk score, the samples were divided into two groups: high-risk group and low-risk group. Patients in the high-risk group suffered a lower survival probability (Fig. 7C). Univariate and multivariate Cox regression analysis suggested that the established model could become an independent predictor, including age, gender, histological grade, stage, TNM stage, and risk score (Fig. 8A, B, Table 3).

Table 2
Identification of differentially expressed DNA methylation sites between Cluster 3 and other Clusters

DNA methylation site	conMean	treatMean	logFC	pValue	fdr
cg00413617	0.325219	0.468693	0.527231	3.79E-07	8.22E-06
cg00765828	0.263837	0.487255	0.88503	6.36E-07	1.32E-05
cg01330280	0.133973	0.092864	-0.52874	0.000205	0.001084
cg01501009	0.295805	0.433979	0.552982	6.88E-07	1.36E-05
cg01760756	0.233642	0.364166	0.640295	6.13E-08	2.15E-06
cg02002583	0.322265	0.515678	0.678223	1.52E-08	9.89E-07
cg02825211	0.298547	0.507008	0.764052	1.16E-08	9.65E-07
cg04524477	0.183084	0.109166	-0.74598	9.77E-05	0.000563
cg04671611	0.304968	0.509565	0.74061	3.36E-08	1.70E-06
cg04682802	0.428694	0.622653	0.538483	3.64E-07	8.22E-06
cg05893785	0.278157	0.195427	-0.50927	5.49E-05	0.000357
cg07098752	0.27403	0.42237	0.624174	3.82E-06	4.82E-05
cg09685934	0.153233	0.242603	0.662869	1.19E-06	1.93E-05
cg10378804	0.149836	0.088401	-0.76125	2.40E-05	0.000195
cg10699171	0.342727	0.216581	-0.66215	0.000302	0.00143
cg11308277	0.204648	0.357149	0.80338	4.74E-08	1.96E-06
cg11718317	0.148513	0.257033	0.791361	3.24E-05	0.000246
cg13614286	0.154288	0.240086	0.63793	1.48E-07	3.97E-06
cg13765961	0.676689	0.470397	-0.52461	9.78E-07	1.71E-05
cg14901205	0.324492	0.590179	0.862969	1.27E-08	9.65E-07
cg15746719	0.330852	0.530793	0.681965	1.10E-06	1.85E-05
cg16508480	0.387702	0.556004	0.520144	4.98E-05	0.000333
cg17102963	0.267475	0.471817	0.818821	1.94E-09	8.56E-07
cg18470295	0.266957	0.417205	0.644148	7.33E-06	7.76E-05
cg18885392	0.492172	0.34046	-0.53168	5.11E-06	5.67E-05
cg21549195	0.356057	0.605258	0.765441	2.06E-07	5.21E-06
cg21609106	0.253179	0.48718	0.944299	9.72E-09	9.65E-07
cg23288973	0.227864	0.4023	0.8201	9.05E-07	1.65E-05
cg24120357	0.094457	0.062855	-0.58764	0.014751	0.028683
cg24464397	0.217307	0.339064	0.641821	8.16E-06	8.44E-05

DNA methylation site	conMean	treatMean	logFC	pValue	fdr
cg24911113	0.482802	0.69979	0.53549	4.35E-08	1.96E-06
cg25563456	0.279144	0.437056	0.646809	1.55E-06	2.44E-05

Table 3
The prognostic effect of different clinical parameters.

Variable	Univariate analysis				Multivariate analysis			
	HR	95%Upper CI	95%Lower CI	P-value	HR	95%Upper CI	95%Lower CI	P-value
age	0.999068	0.958695	1.041141	0.964665	0.989581	0.950541	1.030224	0.610039
gender	0.27971	0.12816	0.61047	0.001377	0.253659	0.097922	0.657081	0.004732
grade	0.859083	0.54693	1.349394	0.509711	0.56247	0.334321	0.946314	0.03017
stage	1.016037	0.637785	1.618617	0.946612	0.612462	0.283796	1.321759	0.2116
T	0.971235	0.669208	1.409572	0.877939	1.720153	0.902914	3.277086	0.099065
M	1.032813	0.72288	1.475629	0.85923	0.644025	0.422328	0.982099	0.040967
N	1.944678	1.456393	2.596671	6.53E-06	2.352073	1.652825	3.347146	2.02E-06
riskScore	2.718282	1.950762	3.78778	3.48E-09	3.094405	2.114331	4.528781	6.13E-09

Discussion

Laryngeal cancer is mainly composed of laryngeal squamous cell carcinoma, which is the second most common malignant type of head and neck tumors. Despite its advantages in diagnosis and treatment, the mortality rate of laryngeal squamous cell carcinoma has remained high in the past 20 years. Epigenetic regulation plays an important role in laryngeal squamous cell carcinoma[21–23]. Gene methylation can inhibit gene expression, and certain histones and DNA methylation are related to laryngeal cancer. Histone deacetylase also plays a role in laryngeal squamous cell carcinoma[24]. According to differences in epigenetic regulation, certain characteristics can be used as diagnostic markers and can even promote the treatment of laryngeal squamous cell carcinoma. More recently, abnormal DNA methylation is one of the hallmarks of cancer tissue. The latest development of sequencing technology makes it possible to analyze the DNA methylation profiles of the whole genome with high resolution. Recently, several promoter sequence-specific methylation related to the occurrence and development of laryngeal cancer has been discovered. The enhancer of zeste homolog 2 (EZH2) is a highly conserved histone methyltransferase, which is overexpressed in different types of cancers such as breast cancer[25] and prostate cancer[26], laryngeal cancer[11]. Lili reported that the occurrence of aberrant methylation events in the LINC00886 TSS may accelerate the malignant progression of laryngeal carcinoma[27]. Another study showed that the regulation of gene expression by TREX2 DNA methylation may affect the incidence and survival of laryngeal cancer[21]. Elevated CMCC3 methylation level is a risk factor for male LSCC patients, especially in patients over 55 years of age who have smoking behavior[28]. Although the influence of some genes with abnormal DNA methylation on the occurrence and development of laryngeal cancer has been widely reported, the comprehensive overview of the classification of laryngeal cancer is based on DNA methylation data still needs further clarification.

Molecular mechanism research based on bioinformatics analysis is an important method in cancer research. Besides, The TCGA database is a publicly available resource that covers a wide variety of data types in a variety of cancers. Therefore, an informatics method was employed to classify the molecular subtype based on DNA methylation profiling. The classification will help to improve the patient's prognosis and seek for a valid target of treatment. Although methylation may serve as an important biomarker in laryngeal carcinoma, the molecular subtypes based on DNA methylation profiles have yet not been reported. We tried to solve the problems in this study by developing a classification method that integrates several DNA methylation biomarkers, which can be used to evaluate the prognosis of treatment effects and help guide treatment choices. Besides, the clinical and statistical significance of these gene methylations concerning tumor classification, survival time, and prognosis needs to be confirmed in the larger cohort. Therefore, specific prognosis subtypes based on DNA methylation status using 112 laryngeal cancer from the TCGA database were explored. Consensus unsupervised clustering was used to classify the molecular subgroups. Unsupervised hierarchical clustering analysis utilizing the Consensus unsupervised clustering method uncovered the five specific DNA methylation subtypes of laryngeal carcinoma. Five subgroups were distinguished by consensus clustering using 11,637 CpGs that significantly influenced survival. Cluster III had a better prognosis compared with Clusters I and II. Moreover, each cluster possessed its own distinct functional enrichment terms. The difference in molecular characteristics in five subtypes would affect the patient's prognosis. Based on the results of the multivariate Cox regression analysis, a prognosis prediction model was constructed.

Nevertheless, there are several limitations to the present study. First, the prognosis prediction model was based on the publicly available dataset, and further clinical studies are needed to confirm these results. Second, it is very difficult to judge the best k in unsupervised consensus analysis. Third, although our research aims to investigate the possibility of constructing a prognostic prediction model, it is still in its infancy and needs improvement.

Based on the TCGA database and a series of bioinformatics methods, five DNA methylation subgroups were identified, and each subgroup has each specific characteristic. At the same time, the specific prognostic methylation sites of patients with laryngeal cancer were determined, and a prognostic prediction model was established, which can help patients with laryngeal cancer identify new biomarkers, precision medicine targets, and disease molecular subtype classification.

Abbreviations

LC: laryngeal carcinoma; CDF: cumulative distribution function; PCA : principal component analysis; SCC: squamous cell carcinomas; KNN: k-nearest neighbors; DEGs: differentially expressed genes.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Data availability could be obtained from the TCGA website.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Author contribution statement:

Y.P.Y; Q.L; J.K.S; W.A wrote the main manuscript text;

Y.P.Y; Q.L; J.K.S prepared Figures 1-8;

J.K.S; W.A contributed to data analysis;

All authors reviewed the manuscript.

Acknowledgment

None.

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. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020.** *CA Cancer J Clin* 2020, **70:**7-30.
3. Megwalu UC, Sikora AG: **Survival outcomes in advanced laryngeal cancer.** *JAMA Otolaryngol Head Neck Surg* 2014, **140:**855-860.
4. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM: **Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.** *Int J Cancer* 2010, **127:**2893-2917.
5. Chen AY, Fedewa S, Zhu J: **Temporal trends in the treatment of early- and advanced-stage laryngeal cancer in the United States, 1985-2007.** *Arch Otolaryngol Head Neck Surg* 2011, **137:**1017-1024.
6. Hinerman RW, Mendenhall WM, Morris CG, Amdur RJ, Werning JW, Villaret DB: **T3 and T4 true vocal cord squamous carcinomas treated with external beam irradiation: a single institution's 35-year experience.** *Am J Clin Oncol* 2007, **30:**181-185.
7. Sher G, Salman NA, Khan AQ, Prabhu KS, Raza A, Kulinski M, Dermime S, Haris M, Junejo K, Uddin S: **Epigenetic and Breast Cancer Therapy: Promising Diagnostic and Therapeutic Applications.** *Semin Cancer Biol* 2020.
8. Zhang C, Zhao N, Zhang X, Xiao J, Li J, Lv D, Zhou W, Li Y, Xu J, Li X: **SurvivalMeth: a web server to investigate the effect of DNA methylation-related functional elements on prognosis.** *Brief Bioinform* 2020.
9. Pandareesh MD, Kameshwar VH, Byrappa KK: **Prostate Carcinogenesis: Insights in relation to Epigenetics and Inflammation.** *Endocr Metab Immune Disord Drug Targets* 2020.
10. Chen R, Zhang M, Zhou Y, Guo W, Yi M, Zhang Z, Ding Y, Wang Y: **The application of histone deacetylases inhibitors in glioblastoma.** *J Exp Clin Cancer Res* 2020, **39:**138.
11. Lian R, Ma H, Wu Z, Zhang G, Jiao L, Miao W, Jin Q, Li R, Chen P, Shi H, Yu W: **EZH2 promotes cell proliferation by regulating the expression of RUNX3 in laryngeal carcinoma.** *Mol Cell Biochem* 2018, **439:**35-43.

12. Shen Z, Chen X, Li Q, Zhou C, Li J, Ye H, Duan S: **SSTR2 promoter hypermethylation is associated with the risk and progression of laryngeal squamous cell carcinoma in males.** *Diagn Pathol* 2016, **11:**10.
13. Pierini S, Jordanov SH, Mitkova AV, Chalakov IJ, Melnicharov MB, Kunev KV, Mitev VI, Kaneva RP, Goranova TE: **Promoter hypermethylation of CDKN2A, MGMT, MLH1, and DAPK genes in laryngeal squamous cell carcinoma and their associations with clinical profiles of the patients.** *Head Neck* 2014, **36:**1103-1108.
14. Gomaa MAM, El Gindy KE, Nabi UGA, Mohammed HM, Twab NAA, Mahmoud R, Mohiy KM: **Human Papillomavirus Subtype 16 and the Pathologic Characteristics of Laryngeal Cancer.** *OTO Open* 2017, **1:**2473974X17707925.
15. Peri S, Izumchenko E, Schubert AD, Slifker MJ, Ruth K, Serebriiskii IG, Guo T, Burtness BA, Mehra R, Ross EA, et al: **NSD1- and NSD2-damaging mutations define a subset of laryngeal tumors with favorable prognosis.** *Nat Commun* 2017, **8:**1772.
16. Torabi SJ, Cheraghlou S, Kasle DA, Savoca EL, Judson BL: **Nonsquamous cell laryngeal cancers: Incidence, demographics, care patterns, and effect of surgery.** *Laryngoscope* 2019, **129:**2496-2505.
17. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM: **The Cancer Genome Atlas Pan-Cancer analysis project.** *Nat Genet* 2013, **45:**1113-1120.
18. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments.** *Bioinformatics* 2012, **28:**882-883.
19. Wilkerson MD, Hayes DN: **ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.** *Bioinformatics* 2010, **26:**1572-1573.
20. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *OMICS* 2012, **16:**284-287.
21. Weigel C, Chaisaingmongkol J, Assenov Y, Kuhmann C, Winkler V, Santi I, Bogatyrova O, Kaucher S, Bermejo JL, Leung SY, et al: **DNA methylation at an enhancer of the three prime repair exonuclease 2 gene (TREX2) is linked to gene expression and survival in laryngeal cancer.** *Clin Epigenetics* 2019, **11:**67.
22. Meng WX, Wang BS: **[Epigenetic regulations in laryngeal squamous cell carcinoma].** *Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi* 2018, **32:**1958-1762.
23. Misawa K, Mochizuki D, Imai A, Mima M, Misawa Y, Mineta H: **Analysis of Site-Specific Methylation of Tumor-Related Genes in Head and Neck Cancer: Potential Utility as Biomarkers for Prognosis.** *Cancers (Basel)* 2018, **10**.
24. Grabarska A, Luszczki JJ, Nowosadzka E, Gumbarewicz E, Jeleniewicz W, Dmoszynska-Graniczka M, Kowalczuk K, Kupisz K, Polberg K, Stepulak A: **Histone Deacetylase Inhibitor SAHA as Potential Targeted Therapy Agent for Larynx Cancer Cells.** *J Cancer* 2017, **8:**19-28.
25. Collett K, Eide GE, Arnes J, Stefansson IM, Eide J, Braaten A, Aas T, Otte AP, Akslen LA: **Expression of enhancer of zeste homologue 2 is significantly associated with increased tumor cell proliferation and is a marker of aggressive breast cancer.** *Clin Cancer Res* 2006, **12:**1168-1174.
26. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, et al: **The polycomb group protein EZH2 is involved in progression of prostate cancer.** *Nature* 2002, **419:**624-629.
27. Lan L, Cao H, Chi W, Meng W, Zhao L, Cui W, Wang B: **Aberrant DNA hypermethylation-silenced LINC00886 gene accelerates malignant progression of laryngeal carcinoma.** *Pathol Res Pract* 2020, **216:**152877.
28. Shen Z, Chen X, Li Q, Zhou C, Xu Y, Yu R, Ye H, Li J, Duan S: **Elevated methylation of CMTM3 promoter in the male laryngeal squamous cell carcinoma patients.** *Clin Biochem* 2016, **49:**1278-1282.

Figures

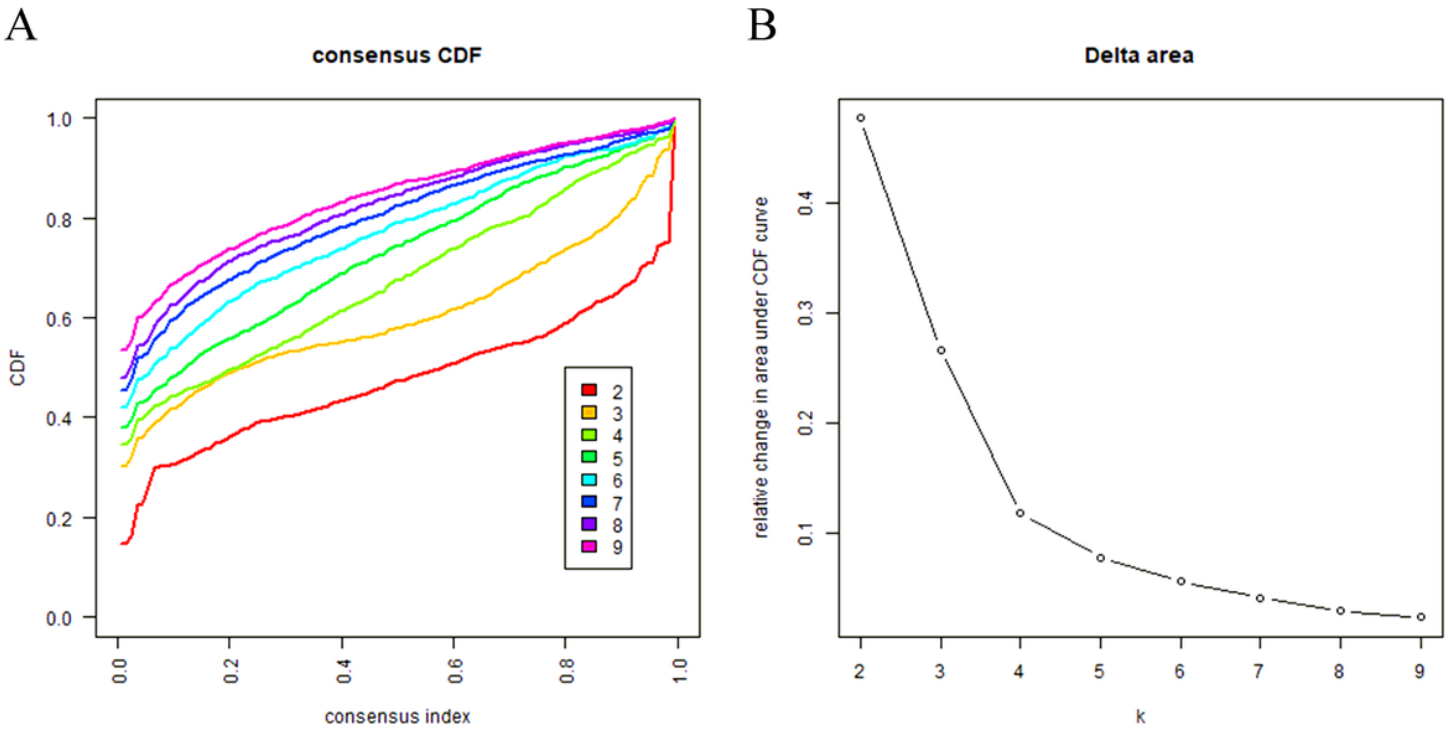

Figure 1

The criteria for selecting the number of clusters. (A) The consensus among clusters of each category number k. (B) The Delta area curve was used for consensus clustering, indicating the relative change of the area under the cumulative distribution function (CDF) curve of each category number k compared to k-1. The horizontal axis represents the category number k, and the vertical axis represents the relative change of the area under the CDF curve.

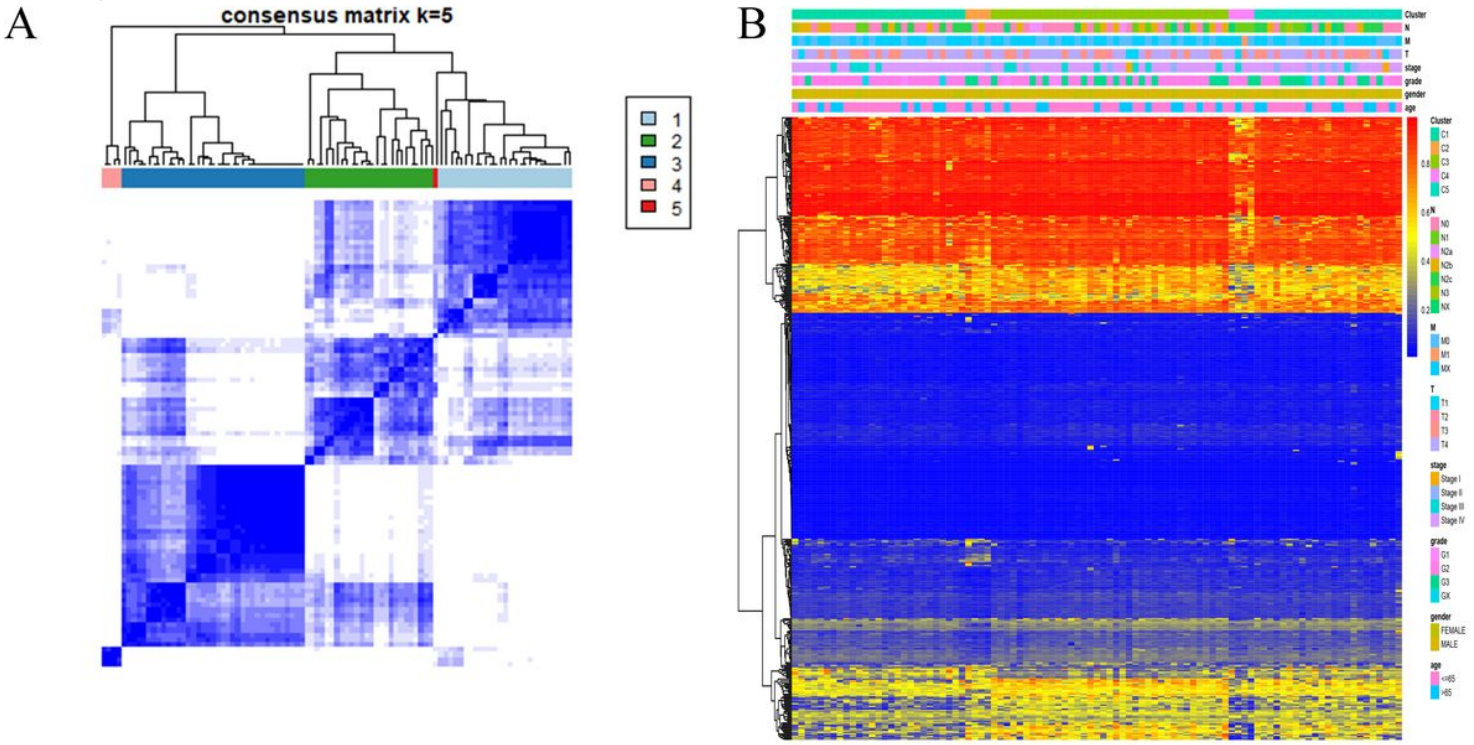

Figure 2

Consensus matrix of DNA methylation classification and corresponding heat map. (A) A color-coded heat map corresponding to the consensus matrix of k = 5 was obtained by applying consensus clustering. The color gradient represents a consensus value from 0 to 1; white corresponds to 0 and dark blue corresponds to 1. (B) Use DNA methylation classification, TNM staging, clinicopathological staging, and histological type as the annotated heat map function to generate a heat map corresponding to the dendrogram in (A).

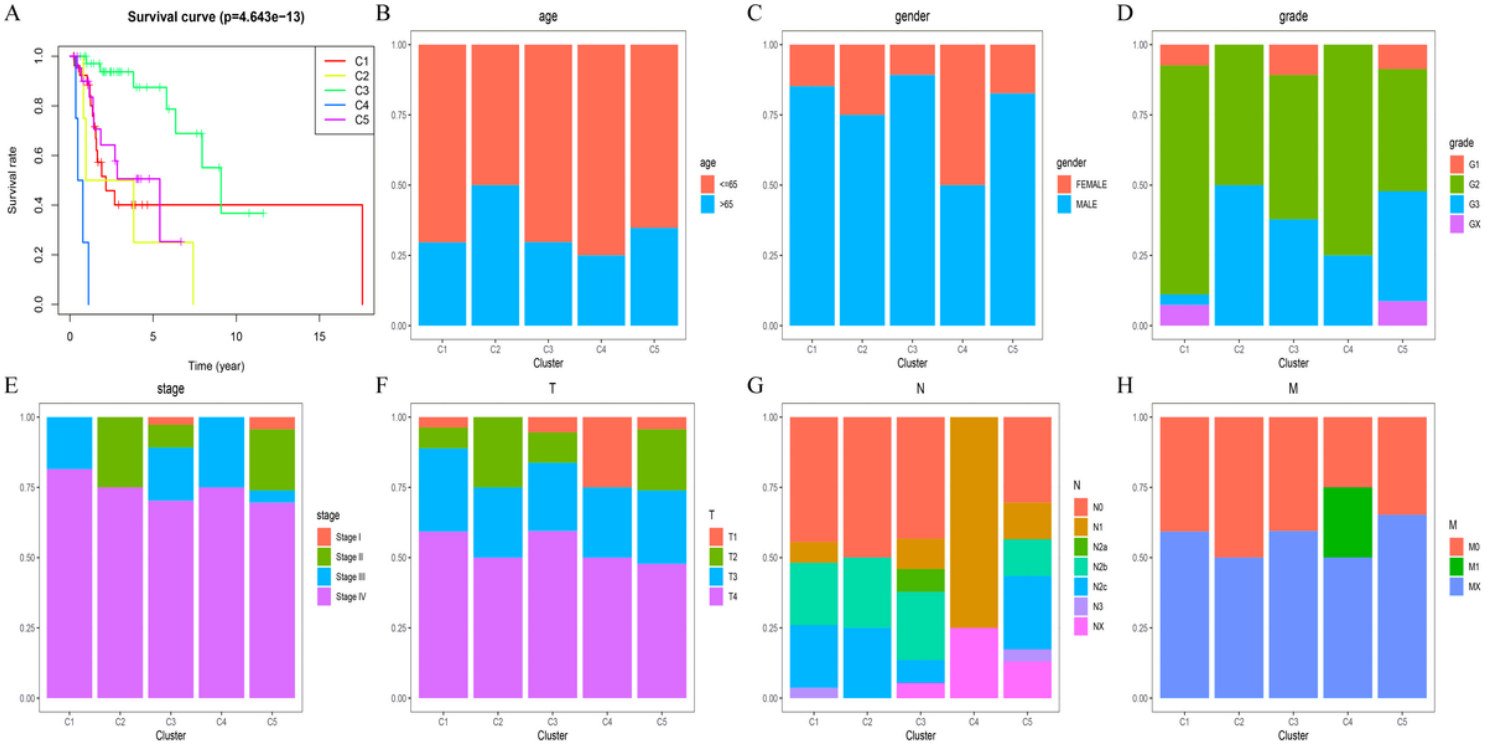

Figure 3

Comparison of prognosis, TNM stage, grade, and age between the DNA methylation clusters. (A) Survival curves for each DNA methylation subtype in the training set. The horizontal axis represents survival time (days), and the vertical axis represents the probability of survival. The number of samples in each cluster is shown in parentheses in the legend. The log-rank test was used to assess the statistical significance of differences between subtypes. Age (B), Gender (C), Grade (D), TNM stage (E), T stage (F), N stage (G), and M stage (H) distributions for each DNA methylation subtype.

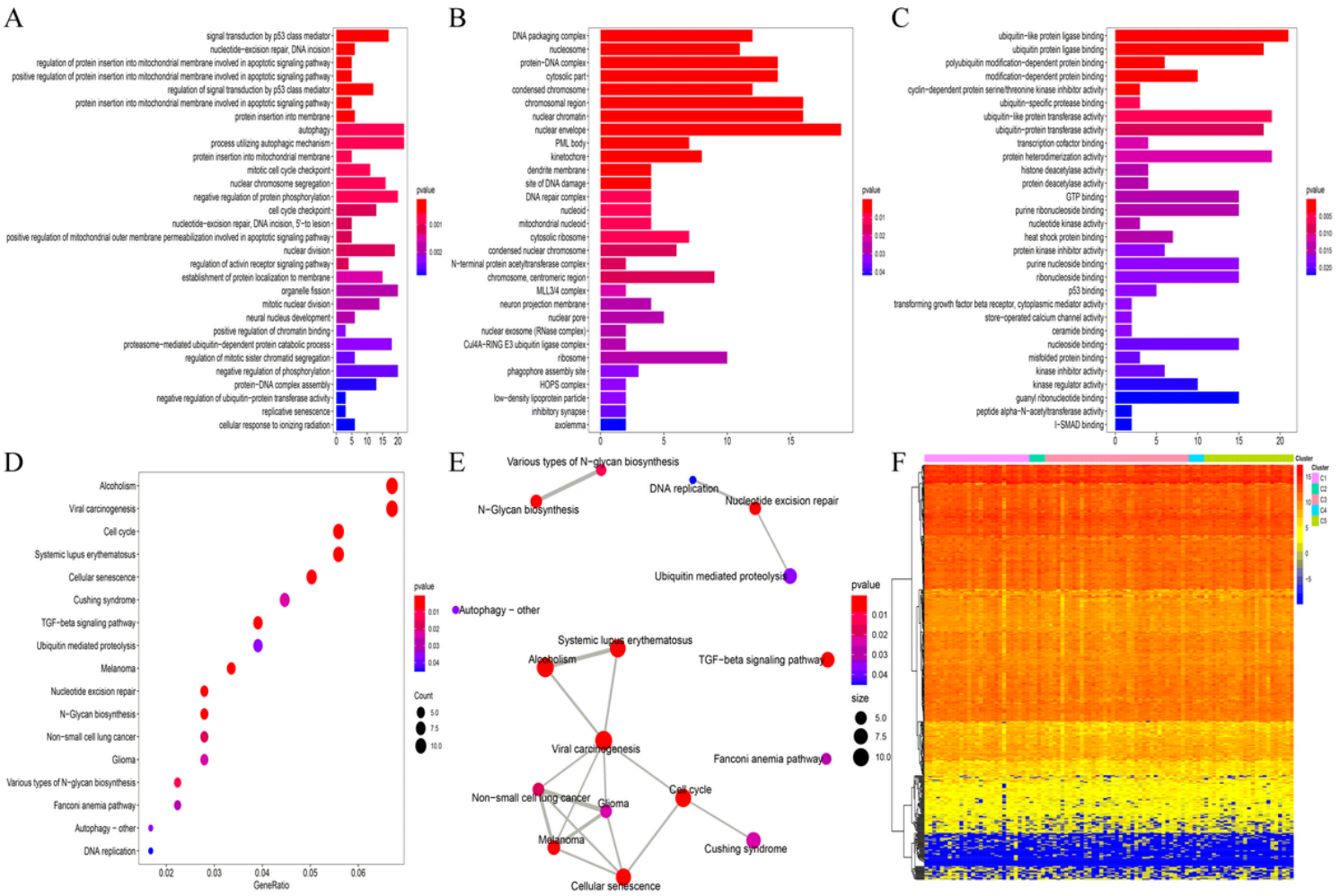

Figure 4

Gene annotations of 455 methylated sites. GO enrichment analysis of annotated genes for the 455 CpG sites, (A)-(C) is for BP, CC, and MF, respectively. (D) KEGG enrichment analysis of annotated genes for the 455 CpG sites. (E) Crosstalk analysis of the enriched KEGG pathways using clusterprofiler package. (F) Cluster analysis heat map for annotated genes associated with the 455 CpG sites.

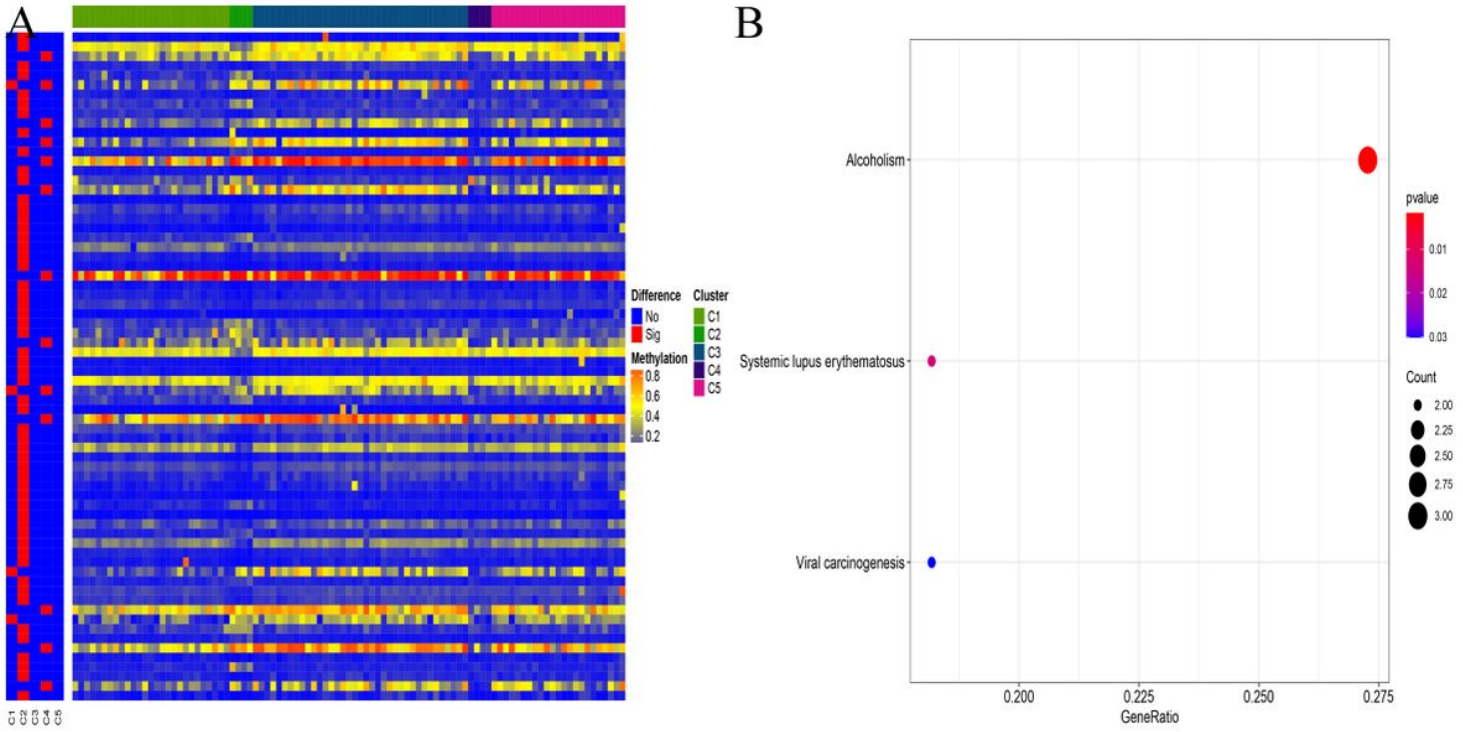

Figure 5

Specific hyper/hypo-methylation CpG sites for each DNA methylation cluster. (A) shows the specific CpG sites for each prognostic subtype of DNA methylation. The red and blue bars represent hypermethylated and hypomethylated CpG sites, respectively. (B) KEGG pathway enrichment analysis of specific CpG sites in cluster 3.

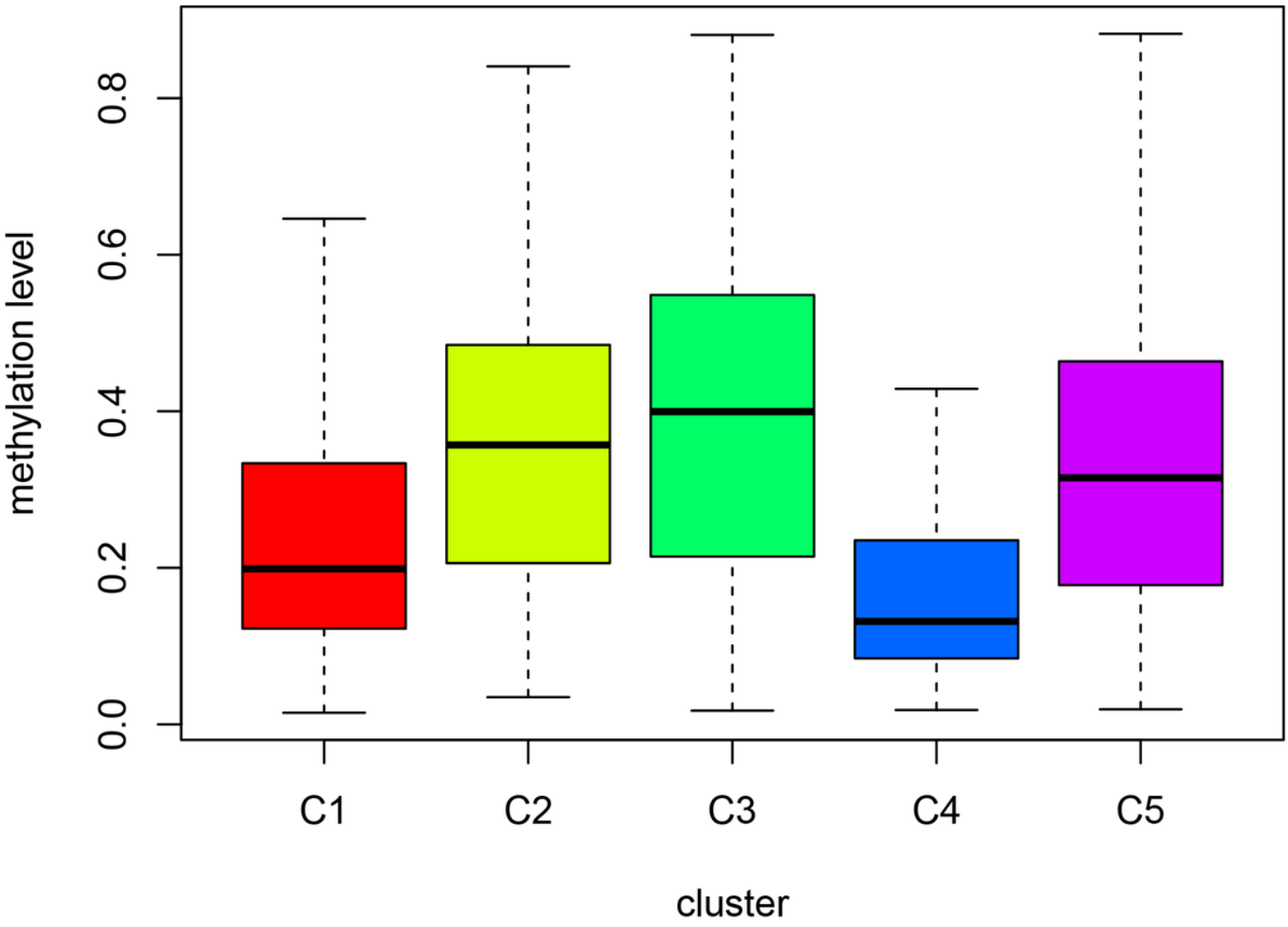

Figure 6

Box plot of CpG methylation levels of the 5 Clusters. Cluster 3 has the highest CpG methylation level.

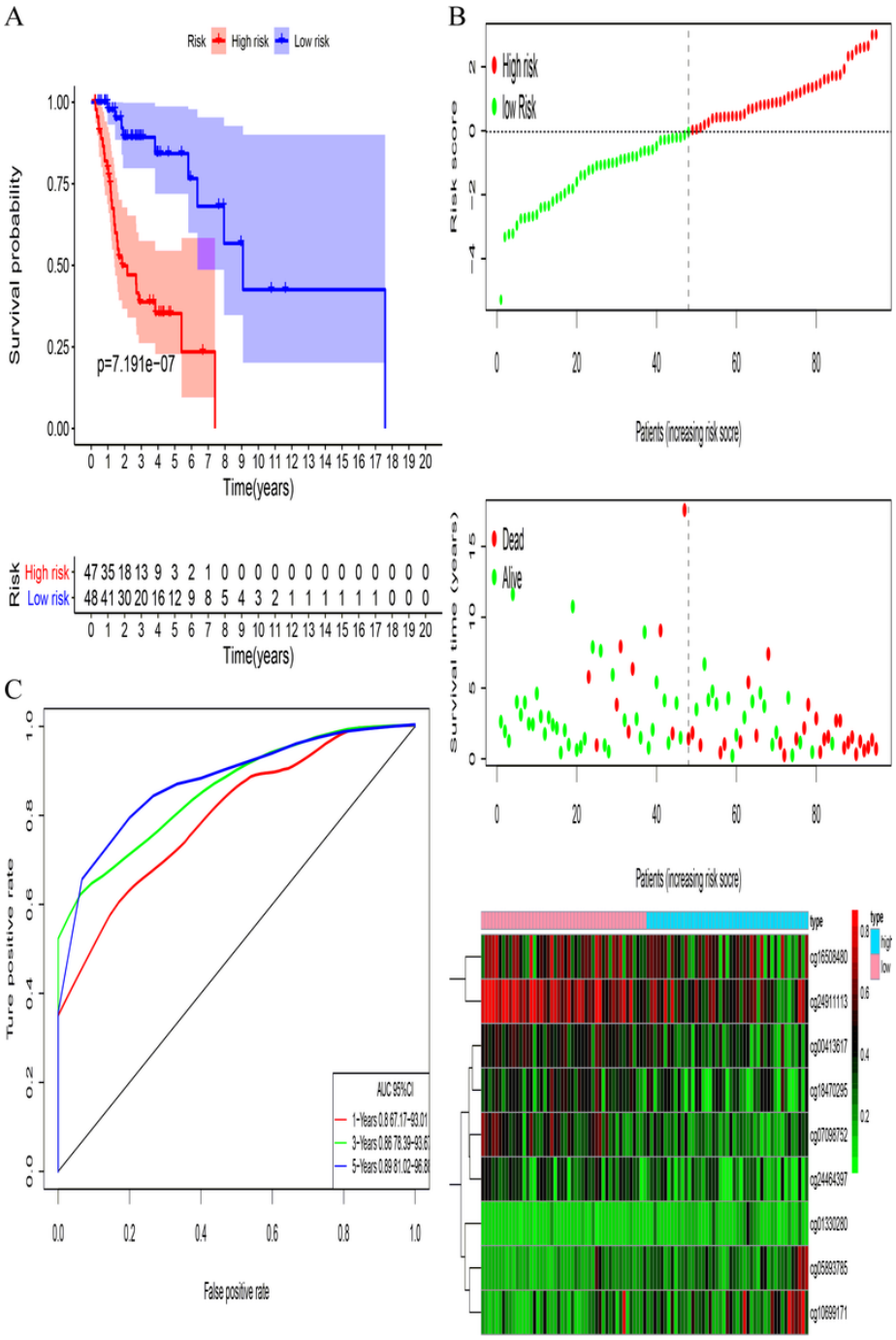

Figure 7

Construction of the prognosis prediction model for laryngeal cancer patients. (A) Analysis of prognostic differences after classification in the TCGA laryngeal cancer dataset. (B) The horizontal axis represents the samples, and the vertical axis represents risk scores (top), overall survival (middle), and methylation site (bottom). (A) Analysis of prognostic differences after classification. (C) ROC curves of prognostic predictors in laryngeal cancer patients.

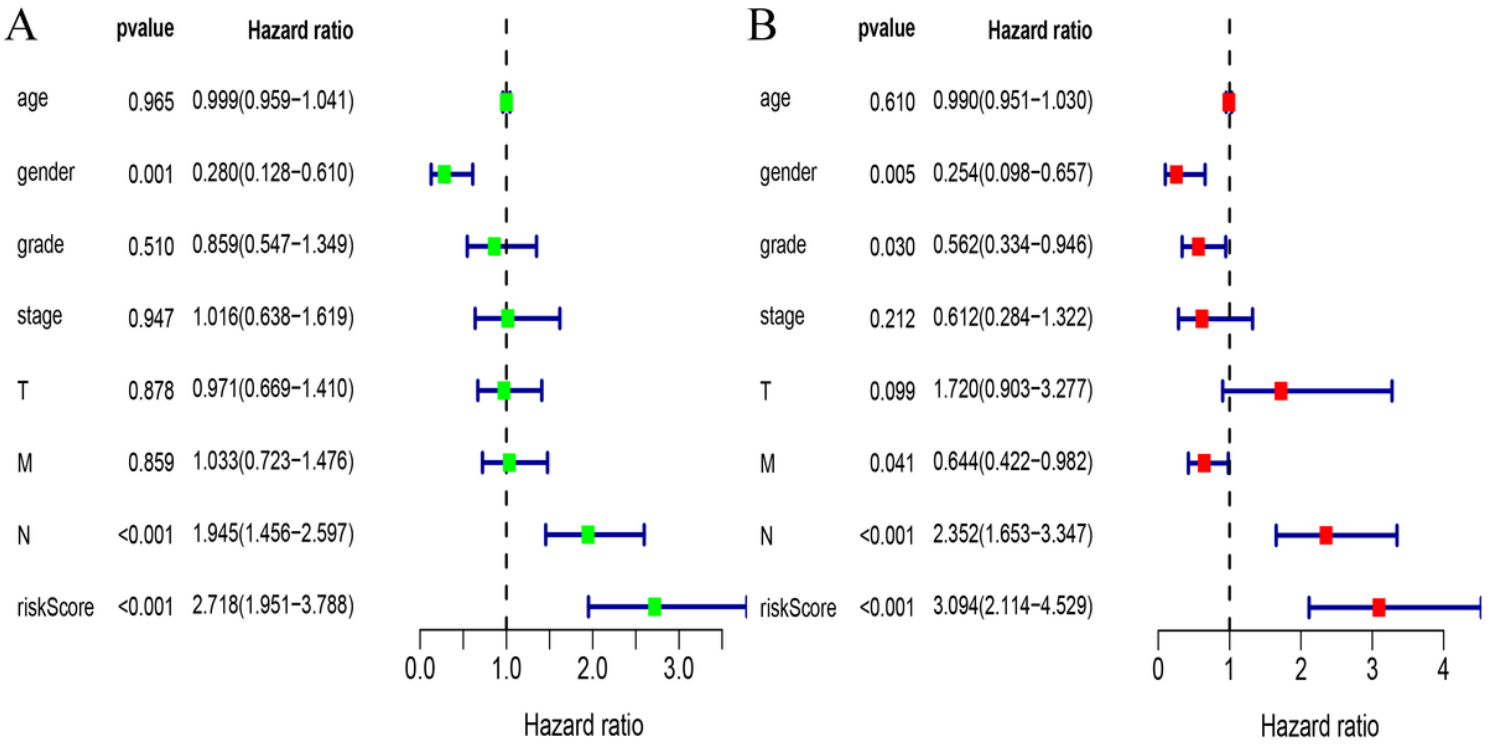

Figure 8

The prediction model could act as an independent predictor in univariate (A) and multivariate regression analysis (B) in the TCGA laryngeal cancer cohort.

Supplementary Files

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

- SupplementaryTable.docx