Development and validation of epigenetic signature predict survival for patients with laryngeal squamous cell carcinoma

Jie Cui
Department of Head and Neck Surgery, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, 510095, Guangdong Province, PR China

Liping Wang
Department of otorhinolaryngology head and neck surgery, The First Affiliated Hospital of Hainan Medical University, Haikou, 570102, Hainan Province, PR China

Waisheng Zhong
Department of Head Neck Surgery, Hunan Cancer Hospital and The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, 410000, Hunan Province, PR China

Zhen Chen
Department of Intensive Care Unit, Shunde Hospital, Southern Medical University (The First people's hospital of Shunde), Foshan, 528308, Guangdong Province, PR China

Jie Chen
Department of Head Neck Surgery, Hunan Cancer Hospital and The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, 410000, Hunan Province, PR China

Hong Yang
Department of Head and Neck Surgery, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, 510095, Guangdong Province, PR China

Genglong Liu (lglong3@mail2.sysu.edu.cn)
Gangzhou Medical Affiliated Cancer Hospital https://orcid.org/0000-0002-9163-4396

Research

Keywords: laryngeal squamous cell carcinoma, DNA methylation-driven genes, epigenetics, prediction, overall survival

Posted Date: April 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-24974/v1

Abstract

Background:

Epigenetic alterations, such as DNA methylation patterns, yield an important role in the initiation, progression and prognosis of laryngeal squamous cell carcinoma (LSCC). We performed a genome-wide integrated analysis of methylation and the transcriptome to establish epigenetic signature to improve the accuracy of survival prediction and optimize therapeutic strategies for LSCC.

Methods:

LSCC DNA methylation datasets and RNA sequencing (RNA-seq) datasets were acquired from the Cancer Genome Atlas (TCGA). MethylMix was applied to detect DNA methylation-driven genes (MDGs). By univariate and multivariate Cox regression analyses, five genes of DNA MDGs was developed an epigenetic signature. The predictive accuracy and clinical value of the epigenetic signature were evaluated by receiver operating characteristic (ROC) and decision curve analysis (DCA), and compared with TNM stage system. Additionally, prognostic value of the epigenetic signature was validated by external Gene Expression Omnibus (GEO) database. According to 5 MDGs of epigenetic signature, the candidate small molecules for LSCC were screen out by the CMap database.

Results:

A total of 88 DNA MDGs were identified, five of which (MAGEB2, SUSD1, ZNF382, ZNF418 and ZNF732) were chosen to construct an epigenetic signature. The epigenetic signature can effectively divide patients into high-risk and low-risk group, with the area under curve (AUC) of 0.8 (5-year OS) and AUC of 0.745 (3-year OS).

Stratification analysis affirmed that the epigenetic signature was still a significant statistical prognostic model in subsets of patients with different clinical variables. Multivariate Cox regression analysis indicated the efficacy of epigenetic signature appears independent of other clinicopathological characteristics. In terms of predictive capacity and clinical usefulness, the epigenetic signature was superior to traditional TNM stage. Additionally, the epigenetic signature was confirmed in external LSCC cohorts from GEO. Finally, CMap matched the 10 most significant small molecules as promising therapeutic drugs to reverse the LSCC gene expression.

Conclusion:

An epigenetic signature, with five DNA MDGs, was identified and validated in LSCC patients by integrating multidimensional genomic data. Compared TNM stage alone, it generates more accurate estimations of the survival probability and maybe offer novel research directions and prospects for individualized treatment of patients with LSCC.

Background

As an aggressively malignant neoplasm, laryngeal squamous cell carcinoma (LSCC) is one of the most prevalent cancers in head and neck region, and represents 85–95% of all laryngeal cancer [1]. According to the American Cancer Society, the estimated respective new cases and new death are 13 150 and 3 710 annually, with incidence and mortality rates of 4.0 and 1.1 per 100 000, respectively [2]. More than 50% of LSCC patients have progressed to the locally advanced stage or metastasis when they are diagnosed due to the absence of the early asymptomatic nature of this cancer, which affects the prognosis of patients [3]. Though treatments have improved during the past decades, long-term survival outcome remains unsatisfactory, with 5-year survival rate only around 60% in the USA [4]. Thus, it is of great importance to ascertain novel biomarkers for valuable molecular targeted therapy and establish predictive models to optimize therapeutic strategies in LSCC patients.

LSCC is a heterogeneous disease in terms of therapeutic response and clinical prognosis. To a certain extent, clinical heterogeneity can be related to distinct molecular subtypes by gene expression patterns [5].RNA expression profiles usually exhibit relative stochastic and rapid variations, which can be directly related to important pathways in malignant cells. DNA methylation, serves as a major epigenetic modification that is involved in the transcriptional regulation of genes and maintains the stability of the genome, is less variable and semi-stable, but shows large variations linked to the activity of cellular processes. Therefore, the combination of transcriptome and epigenetic status would be helpful to identify new markers and improve the accuracy of prognosis. What's more, changes in DNA methylation with a high level of plasticity allows tumor cells to quickly adapt to changes in metabolic restrictions or physiology during the process of tumorigenesis [6, 7]. Hence, it is reasonable to analyses the DNA methylation pattern in the tumor cells to find novel therapeutic targets and predictors for survival in LSCC patients.

The availability of high throughput genomic assays such as RNA-seq and DNA methylation-seq have opened the possibility for a comprehensive characterization of all molecular alterations of cancers, leading to the discovery of new biomarkers of clinical and therapeutic value [8]. In the present study, we performed a genome-wide integrated analysis of methylation and the transcriptome to characterize the crosstalk between DNA methylation and RNA regulation for patients with LSCC in The Cancer Genome Atlas (TCGA) database. We identify methylation-driven genes (MDGs), and then developed an epigenetic signature capable of predicting the overall survival, and further screen correlated small molecule target drugs. The proposed epigenetic signature was validated in external datasets from the GEO database. In addition, we assessed the predictive ability and clinical application of the epigenetic signature and compared it to the TNM stage.

.

Materials And Methods

Sample selection and data processing

We downloaded DNA methylation (111 LSCC samples and 12 normal samples), RNA-sequencing profiles (117 LSCC samples and 16 normal samples) and clinical information data (**Supplementary material 1**) of LSCC patients from TCGA (https://gdc.cancer.gov/), which recorded before February 14, 2020. Excluding the unavailability of DNA methylation or gene transcriptome datasets or overall survival (OS) time < 1 month, as a result, a total of 109 LSCC patients were enrolled in this analysis. The GSE65858 microarray dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65858) includes 270 head and neck squamous cell carcinoma (HNSCC) specimens, 52 specimens of them is LSCC with gene expression profiles and the associated clinical characteristics **(Supplementary material 2)**. The clinical endpoint was overall survival (OS), defined as the time from surgery to death. Besides, patients who were alive at last follow-up are considered as censored observations. All data were normalized in the R computing environment using the edgeR package or Limma package. Methylation data were in the form of β value, representing the ratio of the methylation probe data vs total probe intensities. Then, the average DNA methylation value for all CpG sites correlated with a gene was calculated and merged into a matrix with the function of TCGA-Assembler. Data were utilized according to the data access policy of TCGA and GEO. All analyses were conducted per relevant regulations and guidelines.

Identification of DNA methylation-driven genes

The MethylMix R package was employed for analysis that integrated DNA methylation data for 117 LSCC samples and 16 normal samples and paired gene expression data for 111 LSCC samples to appraise DNA methylation events that have a significant impact on the expression of the corresponding gene, indicating that the gene is a DNA methylation-driven gene (MDG). A total of three procedures of MethylMix analysis were described as previous studies [9].In addition, the differential methylation value where gene with a positive differential methylation value signifies hypermethylation and gene with a negative differential methylation value signifies hypomethylation can be applied in subsequent analysis.

Functional enrichment and pathway analysis of methylation-driven genes

Gene ontology (GO) analysis, including the molecular function (MF), biological process (BP) and cellular component (CC), was performed on identified MDGs using the DAVID database (http://david.abcc.ncifcrf.gov/), which provides integrative and systematic annotation tools for uncovering biological meaning of genes. And we used GOplot R package to visualize the result. Additionally, pathway analysis was carried out for the MDGs with ConsensusPathDB (http://cpdb.molgen.mpg.de/), which is a functional molecular interaction database, integrating information on gene regulation, signal transduction, biochemical metabolism, protein interacting, genetic interacting signaling in humans. Humancyc, Kegg, Reactome, Wikipathways, Smpdb, and Biocarta and Signalink were selected for subsequent analysis. $P < 0.05$ was set as the threshold value.

Construction and verification of an epigenetic signature

First, univariable Cox regression analysis is utilized to select prognostic related MDGs with $P < 0.05$ as the threshold. After primary filtering, the prognostic related MDGs were all assembled into multivariate Cox

regression model, then an epigenetic **s**ignature based on these prognostic related MDGs is developed.

The epigenetic signature risk score was generated through a linear combination of the expression levels of independent DNA MDGs using coefficients from multivariate Cox regression as the weights. On the basis of the median risk score, LSCC patients were classified into two cohorts (high-risk cohorts and low-risk cohorts). Survival differences between high-risk group and low-risk group were assessed by Kaplan-Meier survival analysis and then compared by the log-rank test. Time-dependent receiver operating characteristic (ROC) curves by means of the timeROC package were applied to evaluate predictive performance. Also, stratified analysis base on various clinical characteristics is conducted to evaluate the discrimination ability of epigenetic signature. Importantly, the GSE65858 dataset from the GEO database was applied to validate the predictive value of the epigenetic signature.

Gene set enrichment analysis (GSEA)

We downloaded GSEA software from the GSEA website (http://software.broadinstitute.org/gsea/index.jsp). 109 LSCC patients were categorized into high-risk cohorts and low-risk cohorts, which was served as the phenotypes. Gene sets related to biosignaling on MSigDB (http://software.broad institute.org/gsea/downloads.jsp) could be found on the GSEA home as reference gene sets. Each analysis was repeated 1000 times according to the default weighted enrichment statistical method. Statistically significant pathways were screened based on the false discovery rate (FDR) < 0.25, |enriched score| >0.35, and gene size ≥ 35 as the cutoff criteria.

Independence of the epigenetic signature from clinicopathological features

We carried out univariate and multivariate Cox regression analyses to adjudicate whether the predictive ability of the epigenetic signature may be independent of other clinicopathological characteristics (including smoke history, alcohol history, age, sex, number of positive lymph nodes (LNs), number of LNs, lymph node ratio (LNR), margin status, lymphovascular invasion, histologic grade, TNM stage, N status, T stage, mutation count, fraction genome altered) of LSCC patients in TCGA database.

Additionally, we validate whether the epigenetic signature remains be independent of other clinical features in GSE65858 database (including smoke history, alcohol history, age, sex, pack years, HPV status, treatment modalities, TNM stage, N status, T stage).

Comparison of predictive performance and clinical usefulness between epigenetic signature and TNM stage

ROC analysis by means of the survivalROC package was carried out to investigate and compare the discrimination ability of the epigenetic signature with traditional TNM staging in TCGA and GEO database. Decision curve analysis (DCA) by means of the stdca.R was employed to evaluate the clinical usefulness and net benefit of the epigenetic signature, and compared to TNM stage in TCGA and GEO database [10].

Joint survival analysis and methylated loci associated with OS

To further investigate the impact of prognostic related MDGs on LSCC patients, we carried out joint survival analysis combined methylation and gene expression to identify hub genes associated with prognosis in patients with LSCC by the survival R package. In addition, we retrieved relevant loci for prognostic related MDGs from downloaded LSCC methylation data. We merged the value of corresponding methylated sites into one matrix and conducted univariate Cox analysis to select potential prognosis methylated loci.$P < 0.05$ was regarded as statistically significant.

Identification of candidate small molecules agents

The CMap database (http://www.broadinstitute.org/cmap/), which collects more than 7,000 gene expression profile changes induced by various small molecular drugs, was adopted to investigate candidate small molecules agents for treatment in LSCC patients. 5 prognostic related MDGs of epigenetic signature were divided into down-regulated and up-regulated groups, which were uploaded into CMap database to screen related active small molecules agents. Then, the enrichment scores were calculated, which signify similarity rang from − 1 to 1. A negative connectivity score (closer to − 1) indicated higher similarity between the genes, which represents potential therapeutic value, whereas a positive connectivity score (closer to + 1) demonstrates the matched small molecule can induce the state of LSCC cells.

Statistical analysis

R software (R version 3.5.2) and SPSS statistics 22.0 were utilized to conduct the statistical analysis. A two-sided $P < 0.05$ would be recognized as statistically significant except for where a certain P value has been given.

Results

Identification of MDGs, functional enrichment and pathway analyses

Based on three matrices and steps of MethylMix, in total, 88 genes, 77 hypermethylated genes and 11 hypomethylated genes, were defined as the epigenetic drivers with Cor <−0.3, |logFC| > 0and $P < 0.05$**(Supplementary material 3)**. The GO term enrichment analysis for MDGs shows the top 6 clusters of enriched sets with significant differences (***P*** < 0.05) **(FigureS**1). As to molecular function (MF), MDGs were mainly involved in nucleic acid binding, metal ion binding, and transcription factor activity, sequence-specific DNA binding. For biological processes (BP), MDGs were mainly enriched in regulation of transcription, DNA-templated and transcription, DNA-templated. As to cellular component (CC), MDGs were mainly involved in (CC) intracellular. The results of pathway enrichment analysis indicated that MDGs were most involved in gene expression (transcription), RNA polymerase II transcription and generic transcription pathway **(FigureS**2).

Construction and verification of an epigenetic signature

Utilizing the univariate Cox regression analysis, we identify 5 genes, namely MAGEB2, SUSD1, ZNF382, ZNF418 and ZNF732, associated with OS with ***P*** <0.05 from 88 MDGs **(Supplementary material 4)**. The methylation degree distributions of the 5 prognostic related MDGs are displayed in **Figure1** by methylation β mixed model. **Figure 2** visualized methylation degrees of MAGEB2, SUSD1, ZNF382, ZNF418 and ZNF732 were negatively correlated with respective expressions in LSCC. An epigenetic signature was constructed based on 5 prognostic related MDGs, which were all embedded into multivariate Cox regression model. The epigenetic signature risk score was computed as follows: risk score = (0.0864 expression level of MAGEB2) + (0.4621 expression level of SUSD1) + (-0.2150 expression level of ZNF382) + (-0.0929 expression level of ZNF418) + (-0.0817 expression level of ZNF732). A median cut-off risk score was employed to classify LSCC patients into a high-risk cohorts (n=54) and a low-risk cohorts (n=55) in TCGA database (**Figure3A**). Intuitively, **Figure 3B** the number of deaths was significantly lower in the low-risk cohorts compared with high-risk cohorts. The Kaplan–Meier analysis indicated that in low risk cohorts LSCC patients were more inclined to higher OS time than patients in high risk cohorts (***P***<0.001) (**Figure3C**). Time-dependent receiver operating characteristic (ROC) curves showed that epigenetic signature had a superior prediction capacity, with AUC of 0.80 (5 years OS) and AUC of 0.745(3 years OS) (**Figure 3D**).In addition, stratification analysis was carried out in subsets of patients with different clinical variables (T1-T2 vs T3-T4, N0 vs N1-N3, I-II stage vs III-IV stage, G1-G2 vs G3-G4, positive margin status vs negative margin status, lymphovascular invasion vs no lymphovascular invasion) for epigenetic signature. In T3-T4, N0 or N1-N3, III-IV stage, G1-G2 or G3-G4, negative margin status, lymphovascular invasion or no lymphovascular invasion subgroup, the epigenetic signature was still a statistically and clinically prognostic model (**Figure S3 and S4**). However, in T1-T2, I-II stage and positive margin status subgroup did not reach significant statistics. Noteworthily, external GEO cohorts (GSE65858 database) were utilized to verify the predictive performance of the epigenetic signature. As was displayed in **Figure 4**, patients with low risk score were more prone to survival and had higher OS time than patients with high-risk score, which consistent with the results of the TCGA dataset. Furthermore, the AUC of epigenetic signature (AUC of 5 year OS: 0.756 and AUC 3 year OS: 0.812) confirmed that the predictive accuracy of the prognostic model was satisfactory.

Gene set enrichment analysis

To investigate potential biological pathways for 5 prognostic related MDGs, we carried out the GSEA, and reveal a total of 56 items were significantly enriched with FDR < 0.25. The level of risk score for epigenetic signature was considered as the phenotypes, and the findings uncovered that high-risk level of epigenetic signature may closely correlated with several important crosstalk, comprising of JAK-STAT signaling pathway, mismatch repair, P53 signaling pathway, T cell receptor signaling pathway, B cell receptor signaling pathway, basal transcription factors, cell cycle, ErbB signaling pathway (**Figure 5**).

Independence of the epigenetic signature from clinicopathological features

To investigate whether the epigenetic signature is independent of the clinicopathological characteristics in TCGA database, Univariate Cox regression analysis found that female, the presence of lymphovascular invasion, positive margin status, and high epigenetic signature risk score were associated with shorter OS (**Table 1**). Multivariate Cox regression analysis continued to verify that epigenetic signature was an independent predictor of unfavorable OS (HR: 1.52, 95%CI: 1.13–2.04, ***P***= 0.006), after adjustment for other risk variables. In addition, external GEO cohorts (GSE65858 database) were utilized to validate whether the epigenetic signature is also independent of other clinical features. Univariate and multivariate Cox regression analyses were performed, which indicated that the epigenetic signature a significant independent indicator for OS (HR: 1.88, 95%CI: 1.25–2.83, ***P***= 0.002) by adjusting covariates (**Table 2**).

Comparison of predictive performance and clinical usefulness between epigenetic signature and TNM stage

To evaluate the predictive ability of epigenetic signature, we compared epigenetic signature to AJCC TNM stage model, ROC curve analysis was performed in TCGA database. As was displayed in **Figure 6A and 6B**, the AUC of epigenetic signature for predicting 5- year and 3-year OS were 0.795 and 0.759, respectively, while that of TNM stage model were 0.543 and 0.541, respectively. Similar results were also found in the GEO cohorts. The AUC of epigenetic signature for predicting 5- year and 3-year OS were 0.798 and 0.812, respectively, and the AUC of the TNM stage model were 0.621 and 0.649, respectively (**Figure 6C and 6D**). Finally, the clinical usefulness of the epigenetic signature was measured by the DCA, an abstract statistical concept, which provided visualized information on the clinical value of a model. DCA graphically revealed that the epigenetic signature at diverse cutoff times (5- year and 3-year OS) was superior to the traditional TNM staging based on the continuity of potential death threshold (x-axis) and the net benefit of risk stratification using the model (y-axis) in TCGA cohorts and GEO cohorts (**Figure 7**).

Joint survival analysis and methylated loci associated with OS

Joint survival analysis, that is, the methylation and gene expression matched evaluation, was additionally conducted to exploit the prognostic value of MAGEB2, SUSD1, ZNF382, ZNF418 and ZNF732. The hypermethylation and low-expression of ZNF382 and ZNF418 exhibited a marked correlation with the poor prognosis of patients with LSCC; The hypomethylation and high-expression of MAGEB2 and SUSD1 exhibited a conspicuous association with the unfavorable prognosis of patients with LSCC. Nevertheless, the combination of methylation and expression of ZNF732 did not have a significant effect on the prognosis of LSCC patients. A total of 61 methylation sites (10 methylated sites in MAGEB2, 7 methylated sites in SUSD1, 13 methylated sites in ZNF382, 16 methylated sites in ZNF418, and 15

methylated sites in ZNF732) were screened out and the univariate Cox regression analysis uncovered that 4 key methylation loci (2 specific methylation sites in MAGEB2 and 2 specific methylation sites in ZNF382) were significantly associated with LSCC prognosis (**TableS1**).

Identification of related active small molecules

Table 3 listed the 10 most significant small molecules agents corresponding to gene expression changes of LSCC. Among these small molecules agents, simvastatin (enrichment score = −0.853), thiocolchicoside (enrichment score = −0.888) and parthenolide (enrichment score = −0.846) were related to a highly significant negative fraction and possess potential to reverse the tumor status of LSCC.

Discussion

Integrating TCGA LSCC RNA-seq datasets with DNA methylation datasets by MethylMix tools, we identified 88 DNA MDGs. On the basis of these DNA MDGs, we developed an epigenetic signature, which could accurately categorize patients into high-risk status and low-risk status. Stratification analysis verified that the epigenetic signature remained a significant statistical prognostic model in subsets of patients with different clinical variables. Multivariate Cox regression analysis uncovered the efficacy of epigenetic signature appears independent of other clinicopathological characteristics. With respect to predictive capacity and clinical usefulness, the epigenetic signature was superior to traditional TNM stage. Additionally, the epigenetic signature was validated in external LSCC cohorts from GEO. Finally, CMap matched the 10 most significant small molecules as promising therapeutic drugs to reverse the LSCC gene expression.

An increasing number of researches recognize that epigenetic changes such as the hypermethylation of tumor suppressor genes and hypomethylation of oncogenes in the diagnosis, progression and prognosis of LSCC played a critical role [11–13]. By quantitative methylation-specific polymerase (qMSP) chain reaction assays in 96 LSCC patients, Shen et al [11] revealed that LZTS2 promoter hypermethylation is linked to risk, progression, and prognosis of LSCC,which can serve as a diagnostic and prognostic biomarker for LSCC. Analyzing the LSCC tissues and Hep-2 cells to investigate the methylation status of the CpG islands of MYCT1 and mRNA levels, Yang et al [12] identified that hypermethylation contributed to the transcriptional downregulation of MYCT1 and could inhibit cancer cell differentiation in LSCC. By interfering with its binding to c-Myc, DNA methylation of the CGCG site (− 695 to − 692) of MYCT1 altered the promoter activity in LSCC. Recently, Weigel et al [13] concluded that a regulatory role of TREX2 DNA methylation for gene expression which might be an indicative of incidence and survival in LSCC.TREX2 DNA methylation could be a potential predictor of treatment response and as a biomarker to understand carcinogenesis in stratified epithelia. Additionally, protein levels of altered TREX2 could influence drug-induced DNA damage repair and offer novel tailored therapies in LSCC. These studies suggested the potential clinical implications of DNA methylation in providing novel biomarkers for valuable molecular targeted therapy and establish predictive models to optimize therapeutic strategies in LSCC patients.

Nevertheless, almost all of the studies focused on the methylation status of one gene with limited statistical power in predictive values. Considering the heterogeneity of LSCC, entire molecular signatures derived from high-content genome screens seem to offer better prognostic value.

To our knowledge, this is the first study carried out a genome-wide integrated analysis of methylation and the transcriptome from TCGA database to seek novel biomarker as potential molecular targeted therapy and to create an epigenetic signature for LSCC patients to optimize therapeutic strategies. When applying high-throughput methodology with 450,000 probes, it is necessary to distinguish the epigenetic changes ("driver") that act as effectors of the malignant phenotype from alterations of "passenger" without any biologic function [14]. Hence, a model-based tool (MethylMix) was applied to identify those genes with aberrant methylation and related these data to RNA-seq data affecting gene expression. Consequently, we identified a cohort of 88 MDGs in LSCC. The functional analysis indicated MDGs have mainly attached oneself to gene expression (transcription), RNA polymerase II transcription, transcription factor activity, sequence-specific DNA binding and so on. It was a hint that DNA methylation is involved in the dysregulation of genes with distinct functions and is functionally linked to outcomes in LSCC patients.

On the basis of univariate and multivariate Cox regression analyses, we selected five MDGs (MAGEB2, SUSD1, ZNF382, ZNF418 and ZNF732) to develop an epigenetic signature. It could effectively classify patients into high-risk group with shorter OS and low-risk group with longer OS in TCGA and GEO database. Remarkably, according to stratified analysis, the epigenetic signature was a statistically and clinically prognostic model in T3-T4, N0 or N1-N3, III-IV stage, G1-G2 or G3-G4, negative margin status, lymphovascular invasion or no lymphovascular invasion subgroup. Yet, in T1-T2, I-II stage and positive margin status subgroup did not reach significant statistics. One possible explanation that a small sample size of T1-T2 or I-II stage or positive margin status subgroup, consisting of less than 20 patients in the different subgroup, are not enough to generate an effect of significant statistics. Besides, univariate and multivariate Cox analysis affirmed that epigenetic signature was an independent predictor of unfavorable OS, regardless of other clinicopathologic variables in TCGA and GEO data set. To explore potential biological pathways for 5 prognostic related MDGs, we carried out the GSEA and indicated that 5 MDGs of epigenetic signature was mainly scattered in cancer-related pathways (JAK-STAT signaling pathway, P53 signaling pathway) and immune-related biological processes (T cell receptor signaling pathway, B cell receptor signaling pathway). It implied that 5 MDGs of epigenetic signature maybe involved in initiation,

maintenance, development of LSCC and associated with outcomes in LSCC patients.

Currently, in the context of the global trend toward organ-preserving treatment for LSCC patients (though total laryngectomy helps disease control and improve survival outcome, it has observed adverse effects on patients' quality of life because of permanent tracheostomy, the loss of voice and issues with swallowing), pursuant to the newest National Comprehensive Cancer Network (NCCN) Guidelines, decision making about larynx-preserving are highly relied on tumor TNM staging. In our research, via ROC curve analysis, epigenetic signature shown more precise predictive ability compared with TNM stage

model in TCGA and GEO database, which could effectively identify low-risk patients prone to larynx-preserving treatment and avoid needless total laryngectomy. Interestingly, DCA results indicated that LSCC survival-related treatment decisions based on the epigenetic signature led to more net benefit than treatment decisions based on TNM stage, or treating either all patients or none in TCGA and GEO database. Taken together, the current epigenetic signature would be clinically useful for the clinicians in tailoring survival-associated treatment decisions.

One prominent finding in our study was that combining methylation and RNA expression data with survival analysis, we identified 4 MDGs (MAGEB2, SUSD1, ZNF382, ZNF418), which may be serve as potential biomarkers or drug targets for early diagnosis and prognostic assessment. MAGEB2, as a member of the MAGEB family, which belongs to the cancer testicular antigens, is located in the last exon on chromosome X. The MAGEB2 gene has been reported to be overexpression in several types of tumors, such as esophageal cancer [15] and lung cancer [16], which has been implicated in carcinogenesis and considered as a potential cancer biomarker [17]. Pattini et al. [18] indicated that MAGEB2 is activated by promoter demethylation in HNSCC, which has growth-promoting effects on a minimally transformed oral keratinocyte cell line. SUSD1, which encodes the sushi domain-containing protein 1 precursor, acts as a common motif in protein-protein interaction [19]. SUSD1 presented subthreshold associations with venous thromboembolism was detailed in a report by Tang et al [20]. Nevertheless, information regarding the role of SUSD1 in cancer is lacking.

ZNF382(Zinc finger protein 382), a member of the Krüppel-associated box zinc finger protein family, plays critical role in regulating certain downstream gene expression as a transcription inhibitor. Due to the hypermethylation of promoter, ZNF382 is downregulated, thereby leading to inhibition of gene expression as found in multiple tumors [21, 22]. Functional researchers uncovered that ZNF382 was potential a tumor suppressor, via colony formation, inhibition of cell proliferation, invasion, migration, and acceleration of cell apoptosis as well as tumorigenic potential, in various cancers such as hepatocellular carcinoma (HCC), gastric cancer, esophagus squamous cell carcinoma (ESCC) [23–25]. ZNF418 (zinc finger protein 418), a novel

Krüppel-associated box/Cys2His2 zinc finger protein, which was involved in critical regulators for the initiation, maintenance, and development of tumor via cell differentiation, tumor suppression, proliferation and apoptosis [26, 27]. Wang et al [28] revealed that ZNF418 was identified as a direct target of miR-1204 and mediated the function of miR-1204 in HCC cells, which inhibited HCC progression through MAPK and c-Jun signaling pathways and potentially serves as a new prognostic biomarker and therapeutic target for HCC. Thus, further characterization of molecules such as MAGEB2, SUSD1, ZNF382 and ZNF418 will provide a new perspective for the development and progress of LSCC, and aided to find potential therapeutic targets for LSCC patients.

Another prominent finding in our study was that we identify a set of potential small molecule drugs that mimic the expression of normal cellular genes, analyzing the MDGs in CMap database. Small molecule drugs with a highly significant negative enrichment value possessed the potential to alter the gene

expression of LSCC, and thus inhibiting the progression of tumors. Simvastatin, as a type of statin, specific inhibitors of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, is frequently utilized to treat hypercholesterolemia and prevent cardiovascular diseases. A growing number of researches have reported simvastatin-related potentially beneficial effects in cancers, including prostate cancer, breast cancer, ovarian cancer, and adenocarcinoma [29–32]. Recently, Ma et al [33] revealed that simvastatin induced cell cycle arrest in the G1 phase in C666-1 cells via decreasing the expression of cyclin-dependent kinase 4 (CKD4) and cyclin D1, and increasing p27 expression in C666-1 cells. Simvastatin treatment inhibited protein kinase B and extracellular signal-regulated kinase 1/2 activation, which is a potential chemotherapy agent for nasopharyngeal carcinoma. Thiocolchicoside, a semi-synthetic colchicine derived from plant honeysuckle, is a muscle relaxant and utilized to treat orthopedic disorders and rheumatologic on account of its anti-inflammatory and analgesic mechanisms. Reuter et al [34] demonstrate that thiocolchicoside exerts an effect on anticancer via the NF-κB pathway resulting in inhibition of cyclooxygenase-2 promoter activity and NF-κB reporter activity. However, efficacy and safety of those small molecule drugs (including simvastatin and thiocolchicoside) on LSCC are still not investigated. Hence, it is urgently demanded to verify the effect of these candidate small molecule drugs on treating LSCC in further studies.

Despite the remarkable sense, it is inevitable that limitations also existed in our study. First, we merely extract retrospectively target data (TCGA and GEO datasets) through biological algorithm approaches the absence of fresh clinical samples to screen and verify our results. So, the application of our model remains needed to validate in external and multicenter prospective cohorts with large sample sizes. Second, the MethylMix is an attractive investigative tool to integrate DNA methylation with RNA expression to identify MDGs in cancers. However, the MethylMix focuses on identifying cis-regulatory effects of DNA methylation on gene expression and does not currently model trans-regulatory effects. Further studies are needed to solve the multiple testing challenges on identifying trans-regulatory effects of DNA methylation on gene expression. Third, the 5 MDGs should be further studied and verified to investigate its specific regulatory function and mechanisms in LSCC.

Conclusion

An epigenetic signature, with five DNA MDGs, was identified and validated in LSCC patients by integrating multidimensional genomic data. Compared TNM stage alone, it generates more accurate estimations of the survival probability and maybe help the development of personalized and precise medicine LSCC field.

Abbreviations

LSCC: laryngeal squamous cell carcinoma; RNA-seq: RNA sequencing; TCGA: the Cancer Genome Atlas; MDGs: methylation-driven genes; ROC: receiver operating characteristic; AUC: area under curve; DCA: decision curve analysis; GEO: Gene Expression Omnibus; OS: overall survival; HNSCC: head and neck squamous cell carcinoma; GO: Gene ontology; MF: molecular function, BP: biological process; CC: cellular

component; LNR: lymph node ratio; LNs: lymph nodes; GSEA: Gene set enrichment analysis; qMSP: quantitative methylation-specific polymerase; NCCN: National Comprehensive Cancer Network; ZNF382:Zinc finger protein 382; HCC: hepatocellular carcinoma; ESCC: esophagus squamous cell carcinoma; ZNF418:zinc finger protein 418; HMG-CoA:3-hydroxy-3-methylglutaryl coenzyme A ; CKD4: cyclin-dependent kinase 4

Declarations

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

All the data was obtained from TCGA and GEO, and the informed consent had been attained from the patients before our study.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are provided in supplementary materials and are also made available in the TCGA (https://gdc.cancer.gov/) and GEO (https://www.ncbi.nlm.nih.gov/geo/).

Funding

Not applicable.

Authors' contributions

GLL, JC, and LPW conceived and designed the study. GLL, HY, JC*, WSZ, and ZC drafted the manuscript. GLL, JC, and LPW analyzed and interpreted all the data. GLL, HY, JC*, WSZ, and ZC prepared the figures and tables. GLL, JC, and LPW reviewed and revised the manuscript. All authors have read and approved the manuscript for publication. *: Corresponding author (Jie Chen).

Authors' information

1 Department of Pathology, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, 510095, Guangdong Province, PR China.

2 Department of otorhinolaryngology head and neck surgery, The First Affiliated Hospital of Hainan Medical University, Haikou, 570102, Hainan Province, PR China

3 Department of Head and Neck Surgery, Affiliated Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, 510095, Guangdong Province, PR China.

4 Department of Head Neck Surgery, Hunan Cancer Hospital and The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, 410000, Hunan Province, PR China

5 Department of Intensive Care Unit, Shunde Hospital, Southern Medical University (The First people's hospital of Shunde), Foshan, 528308, Guangdong Province, PR China.

Acknowledgements

Not applicable.

References

1. Unsal AA, Kilic S, Dubal PM, Baredes S, Eloy JA. A population-based comparison of European and North American sinonasal cancer survival. Auris nasus larynx. 2018;45(4):815–24.
2. Siegel RL, Miller KD. **Cancer statistics, 2020**. 2020, 70(1):7–30.
3. Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, Kish JA, Kim HE, Cmelak AJ, Rotman M, et al. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. N Engl J Med. 2004;350(19):1937–44.
4. Hoffman HT, Porter K, Karnell LH, Cooper JS, Weber RS, Langer CJ, Ang KK, Gay G, Stewart A, Robinson RA. Laryngeal cancer in the United States: changes in demographics, patterns of care, and survival. Laryngoscope. 2006;116(9 Pt 2 Suppl 111):1–13.
5. Cui J, Wen Q, Tan X, Chen Z: **A Genomic-Clinicopathologic Nomogram Predicts Survival for Patients with Laryngeal Squamous Cell Carcinoma**. 2019, **2019**:5980567.
6. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, Hilmarsdottir H, Olafsdottir E, Tryggvadottir L, Jonasson JG, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Molecular oncology. 2015;9(3):555–68.
7. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, Fasching PA, Widschwendter M. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. Nature communications. 2016;7:10478.

8. Bao M, Shi R, Zhang K, Zhao Y, Wang Y. **Development of a membrane lipid metabolism-based signature to predict overall survival for personalized medicine in ccRCC patients**. 2019, 10(4):383–393.
9. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, Lin Y, Wang D, Yang X, Zheng Y, et al. DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. Theranostics. 2019;9(24):7251–67.
10. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:i6.
11. Shen Z, Lin L, Cao B, Zhou C, Hao W, Ye D. LZTS2 promoter hypermethylation: a potential biomarker for the diagnosis and prognosis of laryngeal squamous cell carcinoma. World J Surg Oncol. 2018;16(1):42.
12. Yang M, Li W, Liu YY, Fu S, Qiu GB, Sun KL, Fu WN. Promoter hypermethylation-induced transcriptional down-regulation of the gene MYCT1 in laryngeal squamous cell carcinoma. BMC Cancer. 2012;12:219.
13. 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**. 2019, 11(1):67.
14. Cedoz PL, Prunello M, Brennan K, Gevaert O. MethylMix 2.0: an R package for identifying DNA methylation genes. Bioinformatics. 2018;34(17):3044–6.
15. Nagashima H, Sadanaga N, Mashino K, Yamashita K, Inoue H, Mori M, Sugimachi K. Expression of MAGE-B genes in esophageal squamous cell carcinoma. Japanese journal of cancer research: Gann. 2001;92(2):167–73.
16. Jang SJ, Soria JC, Wang L, Hassan KA, Morice RC, Walsh GL, Hong WK, Mao L. Activation of melanoma antigen tumor antigens occurs early in lung carcinogenesis. Cancer research. 2001;61(21):7959–63.
17. Almstedt M, Blagitko-Dorfs N, Duque-Afonso J, Karbach J, Pfeifer D, Jager E, Lubbert M. The DNA demethylating agent 5-aza-2'-deoxycytidine induces expression of NY-ESO-1 and other cancer/testis antigens in myeloid leukemia cells. Leukemia research. 2010;34(7):899–905.
18. Pattani KM, Soudry E, Glazer CA, Ochs MF, Wang H, Schussel J, Sun W, Hennessey P, Mydlarz W, Loyo M, et al. MAGEB2 is activated by promoter demethylation in head and neck squamous cell carcinoma. PloS one. 2012;7(9):e45534.
19. Komaromi I, Bagoly Z, Muszbek L. Factor XIII: novel structural and functional aspects. Journal of thrombosis haemostasis: JTH. 2011;9(1):9–20.
20. Tang W, Teichert M, Chasman DI, Heit JA, Morange PE, Li G, Pankratz N, Leebeek FW, Pare G, de Andrade M, et al. A genome-wide association study for venous thromboembolism: the extended cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. Genet Epidemiol. 2013;37(5):512–21.

21. Hudson NO, Buck-Koehntop BA. **Zinc Finger Readers of Methylated DNA**. *Molecules (Basel, Switzerland)* 2018, 23(10).
22. Cheng Y, Geng H, Cheng SH, Liang P, Bai Y, Li J, Srivastava G, Ng MH, Fukagawa T, Wu X, et al. KRAB zinc finger protein ZNF382 is a proapoptotic tumor suppressor that represses multiple oncogenes and is commonly silenced in multiple carcinomas. Cancer research. 2010;70(16):6516–26.
23. Dang S, Zhou J, Chen Y, Chen P, Ji M, Shi B, Yang Q, Hou P. Dynamic expression of ZNF382 and its tumor-suppressor role in hepatitis B virus-related hepatocellular carcinogenesis. Oncogene. 2019;38(24):4804–19.
24. Pei L, He X, Li S, Sun R, Xiang Q, Ren G, Xiang T. KRAB zinc-finger protein 382 regulates epithelial-mesenchymal transition and functions as a tumor suppressor, but is silenced by CpG methylation in gastric cancer. Int J Oncol. 2018;53(3):961–72.
25. Zhang C, Xiang T, Li S, Ye L, Feng Y, Pei L, Li L, Wang X, Sun R, Ren G, et al. The novel 19q13 KRAB zinc-finger tumour suppressor ZNF382 is frequently methylated in oesophageal squamous cell carcinoma and antagonises Wnt/beta-catenin signalling. Cell death disease. 2018;9(5):573.
26. Pu W, Wang C, Chen S, Zhao D, Zhou Y, Ma Y, Wang Y, Li C, Huang Z, Jin L, et al. Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC). Clinical epigenetics. 2017;9:129.
27. Hui HX, Hu ZW, Jiang C, Wu J, Gao Y, Wang XW. ZNF418 overexpression protects against gastric carcinoma and prompts a good prognosis. OncoTargets therapy. 2018;11:2763–70.
28. Wang L, Sun L, Wang Y, Yao B, Liu R, Chen T, Tu K, Liu Q, Liu Z. miR-1204 promotes hepatocellular carcinoma progression through activating MAPK and c-Jun/AP1 signaling by targeting ZNF418. Int J Biol Sci. 2019;15(7):1514–22.
29. Parikh A, Childress C, Deitrick K, Lin Q, Rukstalis D, Yang W. Statin-induced autophagy by inhibition of geranylgeranyl biosynthesis in prostate cancer PC3 cells. Prostate. 2010;70(9):971–81.
30. Feldt M, Bjarnadottir O, Kimbung S, Jirstrom K, Bendahl PO, Veerla S, Grabau D, Hedenfalk I, Borgquist S. Statin-induced anti-proliferative effects via cyclin D1 and p27 in a window-of-opportunity breast cancer trial. Journal of translational medicine. 2015;13:133.
31. Liu H, Liang SL, Kumar S, Weyman CM, Liu W, Zhou A. Statins induce apoptosis in ovarian cancer cells through activation of JNK and enhancement of Bim expression. Cancer Chemother Pharmacol. 2009;63(6):997–1005.
32. Ogunwobi OO, Beales IL. Statins inhibit proliferation and induce apoptosis in Barrett's esophageal adenocarcinoma cells. Am J Gastroenterol. 2008;103(4):825–37.
33. Ma Z, Wang W, Zhang Y, Yao M, Ying L, Zhu L. Inhibitory effect of simvastatin in nasopharyngeal carcinoma cells. Experimental therapeutic medicine. 2019;17(6):4477–84.
34. Reuter S, Prasad S, Phromnoi K, Ravindran J, Sung B, Yadav VR, Kannappan R, Chaturvedi MM, Aggarwal BB. Thiocolchicoside exhibits anticancer effects through downregulation of NF-kappaB pathway and its regulated gene products linked to inflammation and cancer. Cancer prevention research (Philadelphia Pa). 2010;3(11):1462–72.

Supplementary File Legends

TableS1. Screening of 4 prognostic risk loci associated with MDGs in LSCC.

FigureS1. Gene ontology analysis of 88 MDGs of LSCC

FigureS2. Functional pathway analysis for 88 MDGs based on ConsensusPathDB database. Only the pathways which $P < 0.05$ were shown here. Node size: the number of genes; Node color: P-value; Edge width: percentage of shared genes; Edge color: genes from input.

FigureS3. Kaplan–Meier survival analysis according to the epigenetic signature stratified by clinicopathological factors. (A) T stage- T1 to T2, (B) T stage- T3 to T4; (C) lymph node status-node negative, (D) lymph node status-node positive; (E) TNM stage- stage I-II, (F) TNM stage- stage III-IV.

FigureS4. Kaplan–Meier survival analysis according to the epigenetic signature stratified by clinicopathological factors. (A) Tumor grade - G1 to G2, (B) Tumor grade - G3 to G4; (C) margin status-negative, (D) margin status - positive; (E) lymphovascular invasion-no, (F) lymphovascular invasion-yes.

FigureS5. Kaplan–Meier survival curves for the joint survival analysis-the combination of gene ZNF732 methylation and expression.

Tables

Table1. Univariable and multivariable Cox regression analysis for prediction of OS in TCGA database.

Factors	Subgroup	Univariable analysis		Multivariable analysis	
		HR(95%CI)	*P*	HR(95%CI)	*P*
Age		1.00 (0.97-1.04)	0.984	NA	NA
Sex	Female	1			
	Male	0.28(0.14-0.56)	0.000*	0.31(0.11-0.87)	0.026*
Smoking history	No	1			
	Yes	0.65(0.35-1.18)	0.156	NA	NA
Alcohol history	No	1			
	Yes	0.77 (0.43-1.39)	0.388	NA	NA
ber of Lymph nodes		1.00 (0.98-1.01)	0.558	NA	NA
ıber of positive LNs		1.00 (0.95-1.04)	0.892	NA	NA
ymph node ratio		1.41 (0.28-7.13)	0.675	NA	NA
Margin status	Negative	1		1	
	Positive	4.68(2.08-10.51)	0.000*	4.29 (1.62-11.33)	0.003*
Lymphovascular invasion	No	1		1	
	Yes	4.10(2.23-7.55)	0.000*	2.06 (0.90-4.68)	0.086
Tumor grade	G1-G2	1			
	G3-G4	0.51(0.25-1.04)	0.064	NA	NA
Clinical T	T1-T2	1			
	T3-T4	0.72(0.35-1.50)	0.376	NA	NA
Clinical N	N0	1			
	N1-N3	1.44(0.80-2.57)	0.222	NA	NA
Clinical stage	I-II	1			
	III-IV	0.86(0.36-2.03)	0.729	NA	NA
Mutation count		0.99(0.98-1.01)	0.542	NA	NA
ion Genome altered		1.44(0.29-7.18)	0.654	NA	NA
igenetic signature		1.64(1.34-2.01)	0.000*	1.52(1.13-2.04)	0.006*

Abbreviations: HR=hazard ratio, CI =confidence intervals, OS=overall survival

NOTE: NA, not available. These variables were eliminated in the multivariate cox regression model, so the HR and ***P*** values were not available. ***P*** < 0.05

Table2. Univariable and multivariable Cox regression analysis for prediction of OS in GEO database.

Factors	Subgroup	Univariable analysis		Multivariable analysis	
		HR(95%CI)	*P*	HR(95%CI)	*P*
Age		0.92(0.86-0.99)	0.024*	0.95(0.90-1.01)	0.105
Sex	Female	1			
	Male	0.28(0.14-0.56)	0.000*	NA	NA
ıoking history	No	1			
	Yes	0.29(0.02-3.92)	0.348	NA	NA
cohol history	No	1			
	Yes	0.74 (0.12-4.55)	0.743	NA	NA
Pack years		1.02 (0.97-1.06)	0.457	NA	NA
HPV status	Negative	1			
	Positive	0.12(0.08-35.24)	0.933	NA	NA
Clinical T	T1-T2	1			
	T3-T4	1.12(0.36-3.52)	0.952	NA	NA
Clinical N	N0	1			
	N1-N3	1.31(0.80-2.12)	0.514	NA	NA
linical stage	I-II	1			
	III-IV	1.06(0.72-2.54)	0.952	NA	NA
Treatment modalities	Mono-treatment	0.51(0.10-2.59)	0.417	NA	NA
	Mutil-treatment	1.44(0.29-7.18)	0.654	NA	NA
enetic signature		1.46(1.16-1.85)	0.002*	1.88(1.25-2.83)	0.002*

Abbreviations: HR=hazard ratio, CI =confidence intervals, OS=overall survival

NOTE: NA, not available. These variables were eliminated in the multivariate cox regression model, so the HR and ***P*** values were not available.***P*** < 0.05

Table3. List of the 10 most significant small molecule drugs that can reverse the tumoral status of LSCC.

CMap name	Mean	N	Enrichment	*P*	Specificity	Percent non-null
thiocolchicoside	-0.824	4	-0.888	0.00036	0	100
liothyronine	-0.796	4	-0.796	0.00338	0.0052	100
simvastatin	-0.779	4	-0.853	0.00092	0	100
parthenolide	-0.762	4	-0.846	0.00101	0.0276	100
medrysone	-0.75	6	-0.787	0.00018	0.0108	100
metamizole sodium	-0.683	6	-0.695	0.00199	0.0526	83
sulfafurazole	-0.649	5	-0.69	0.00653	0.0341	80
norcyclobenzaprine	-0.643	4	-0.736	0.00969	0.0388	100
citiolone	-0.641	6	-0.648	0.0052	0.0056	83
lanatoside C	-0.529	6	-0.63	0.00793	0.1277	83

Figures

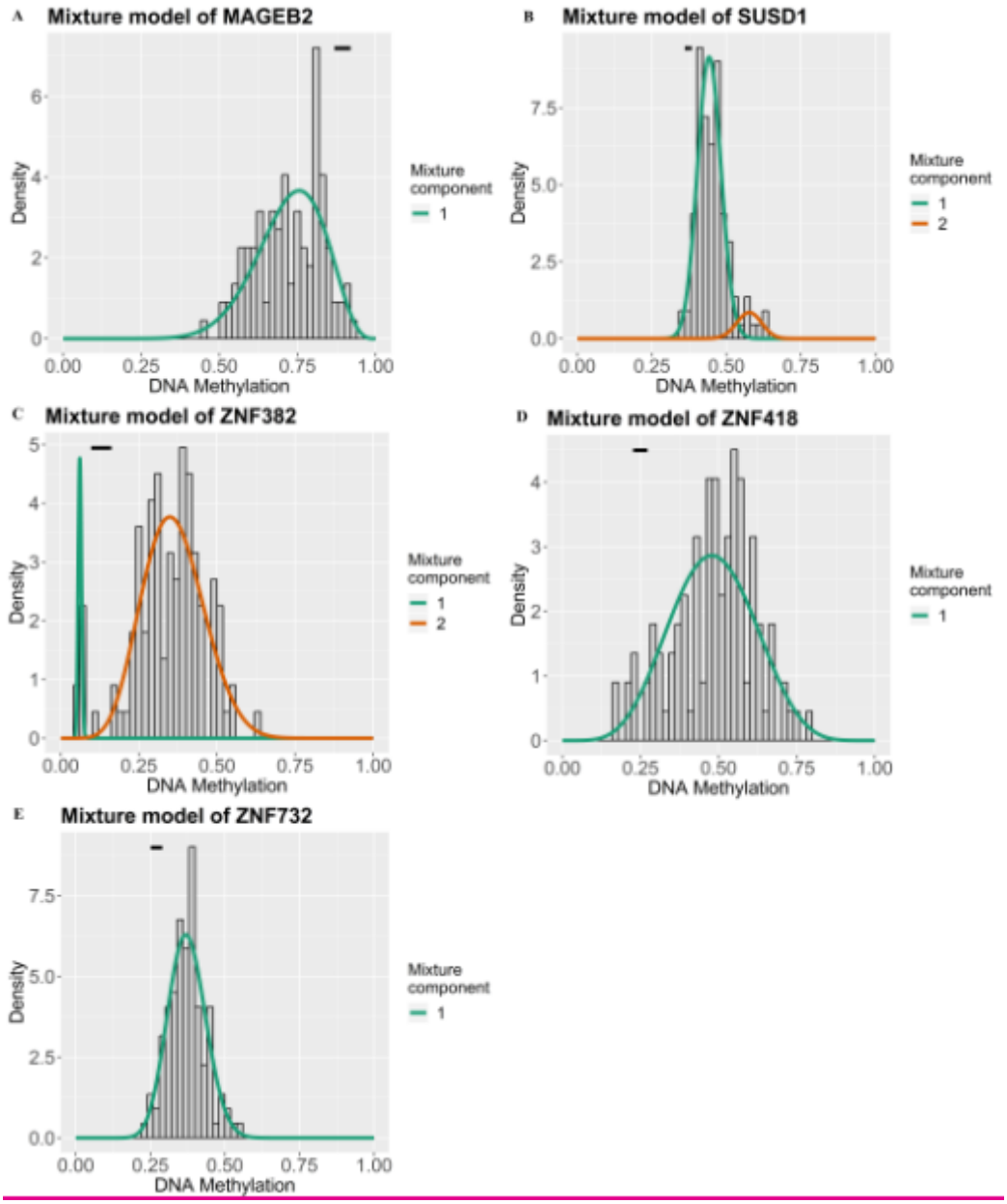

Figure 2

Distribution map of the methylation degree of the five MDGs in the epigenetic signature. The X-axis indicates the degree of methylation, the Y-axis indicates the number of methylated samples, and the histogram shows the subgroups of the methylation distribution in the cancer group. The black horizontal line indicates the methylation degree distribution of the normal samples. The simulated trend of methylation distribution in cancer tissues is represented by the curve.

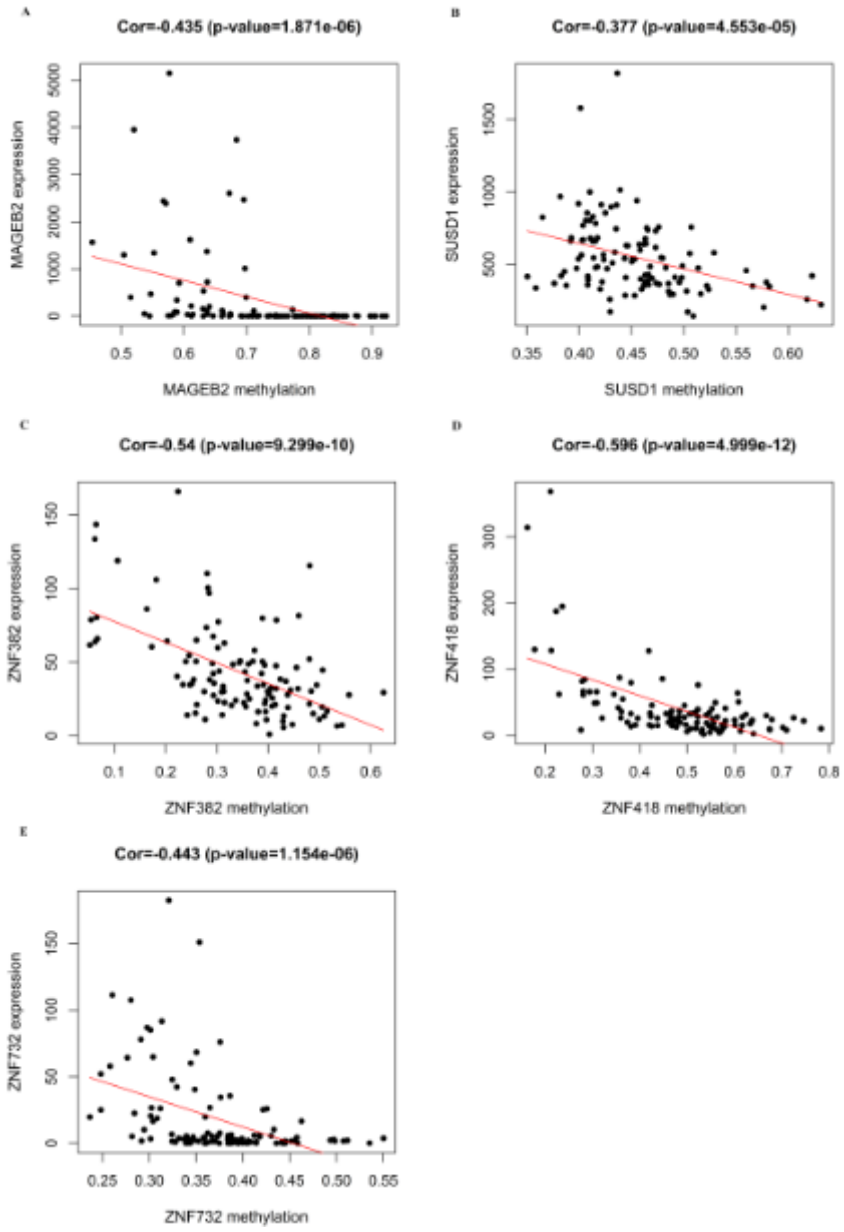

Figure 4

Correlation between the methylation degree and expression of the five MDGs in the epigenetic signature. The X-axis indicates the gene methylation degree β value, and the Y-axis indicates the gene expression level

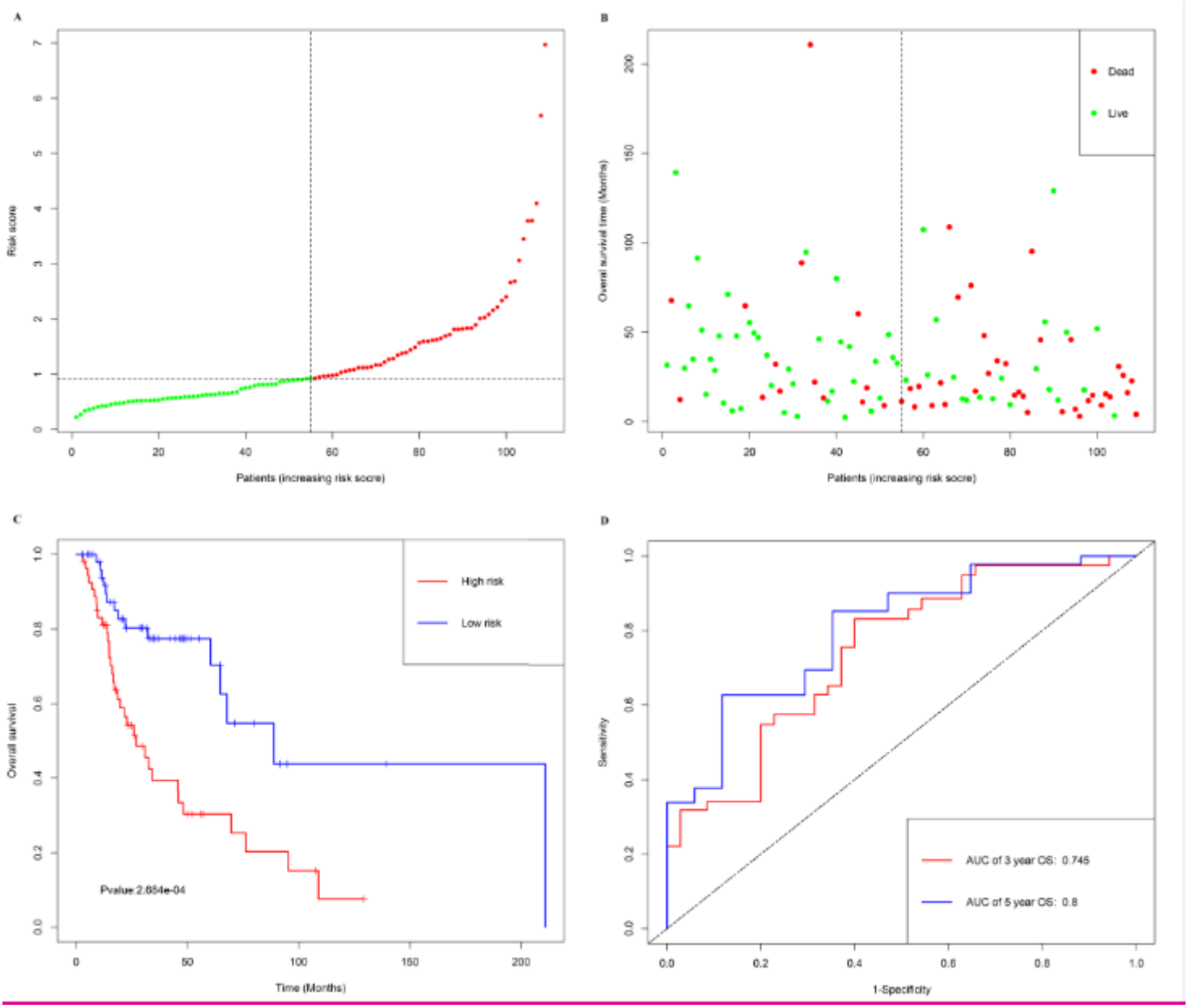

Figure 6

Development of epigenetic signature for prediction of survival in LSCC patients in TCGA database. (A and B) Distribution of epigenetic signature risk score. (C) Time-independent ROC curves with AUC values to evaluate predictive efficacy of epigenetic signature risk score. (D) Kaplan-Meier estimates of patients' survival status and time using the median risk score cut-off which divided patients into low-risk and high-risk groups.

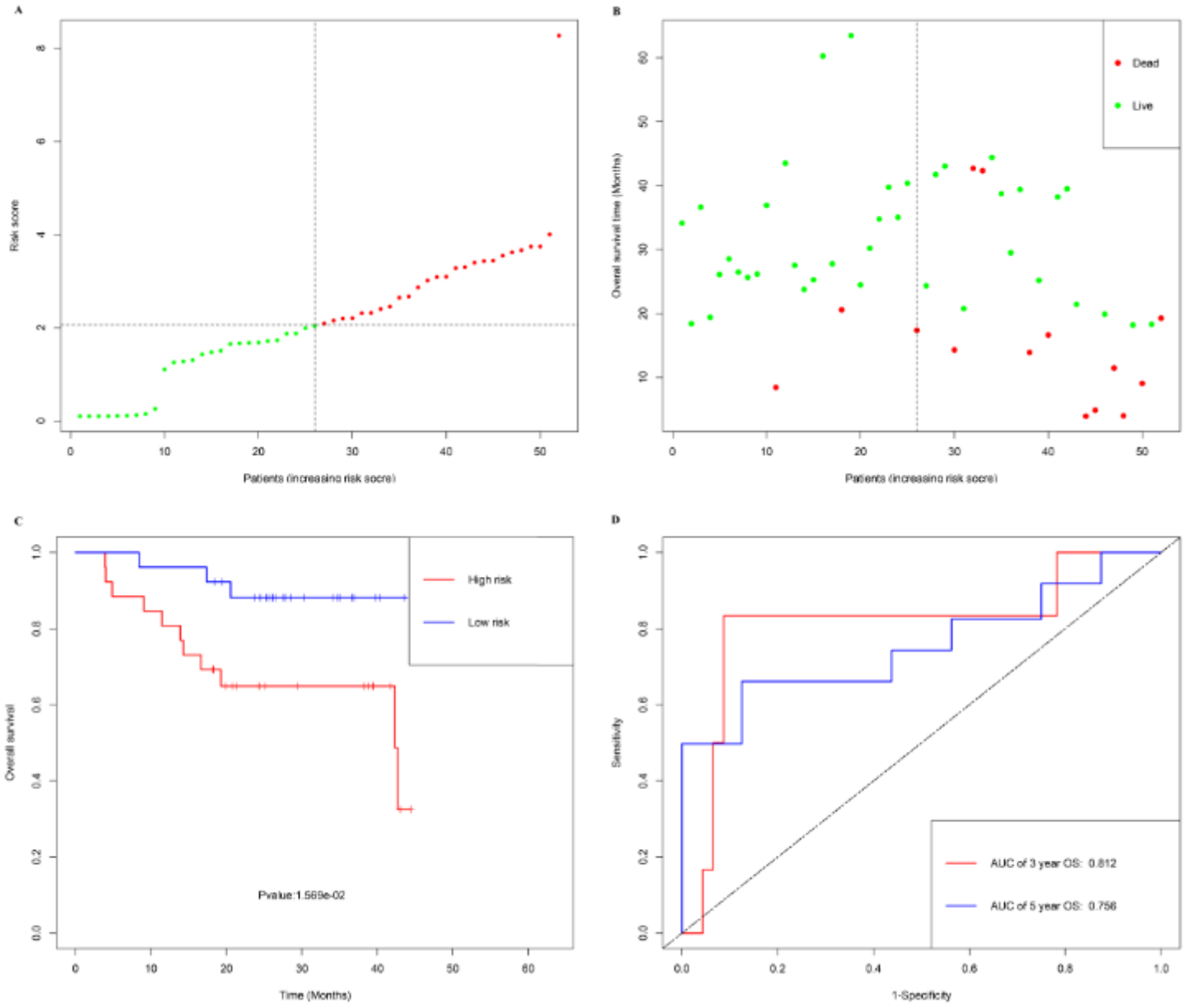

Figure 8

Development of epigenetic signature for prediction of survival in LSCC patients in GEO database. (A and B) Distribution of epigenetic signature risk score. (C) Time-independent ROC curves with AUC values to evaluate predictive efficacy of epigenetic signature risk score. (D) Kaplan-Meier estimates of patients' survival status and time using the median risk score cut-off which divided patients into low-risk and high-risk groups.

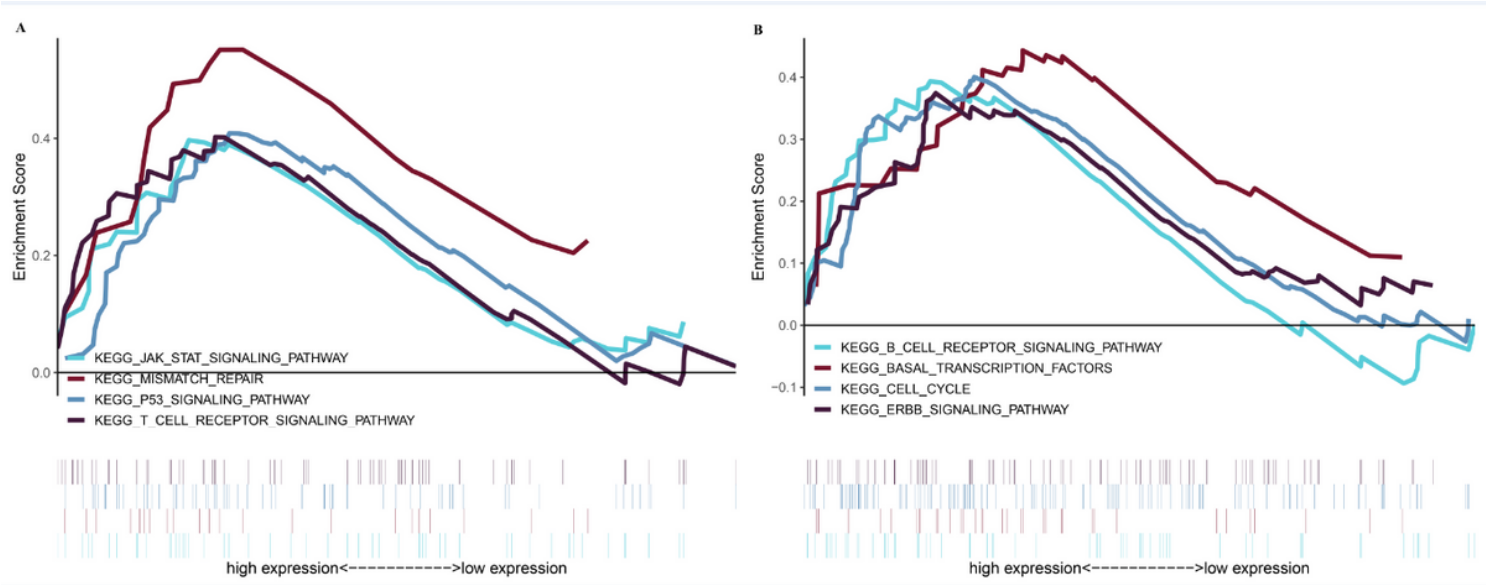

Figure 10

Gene set enrichment analysis for identification of the underlying pathways using risk score as the phenotype (A and B).

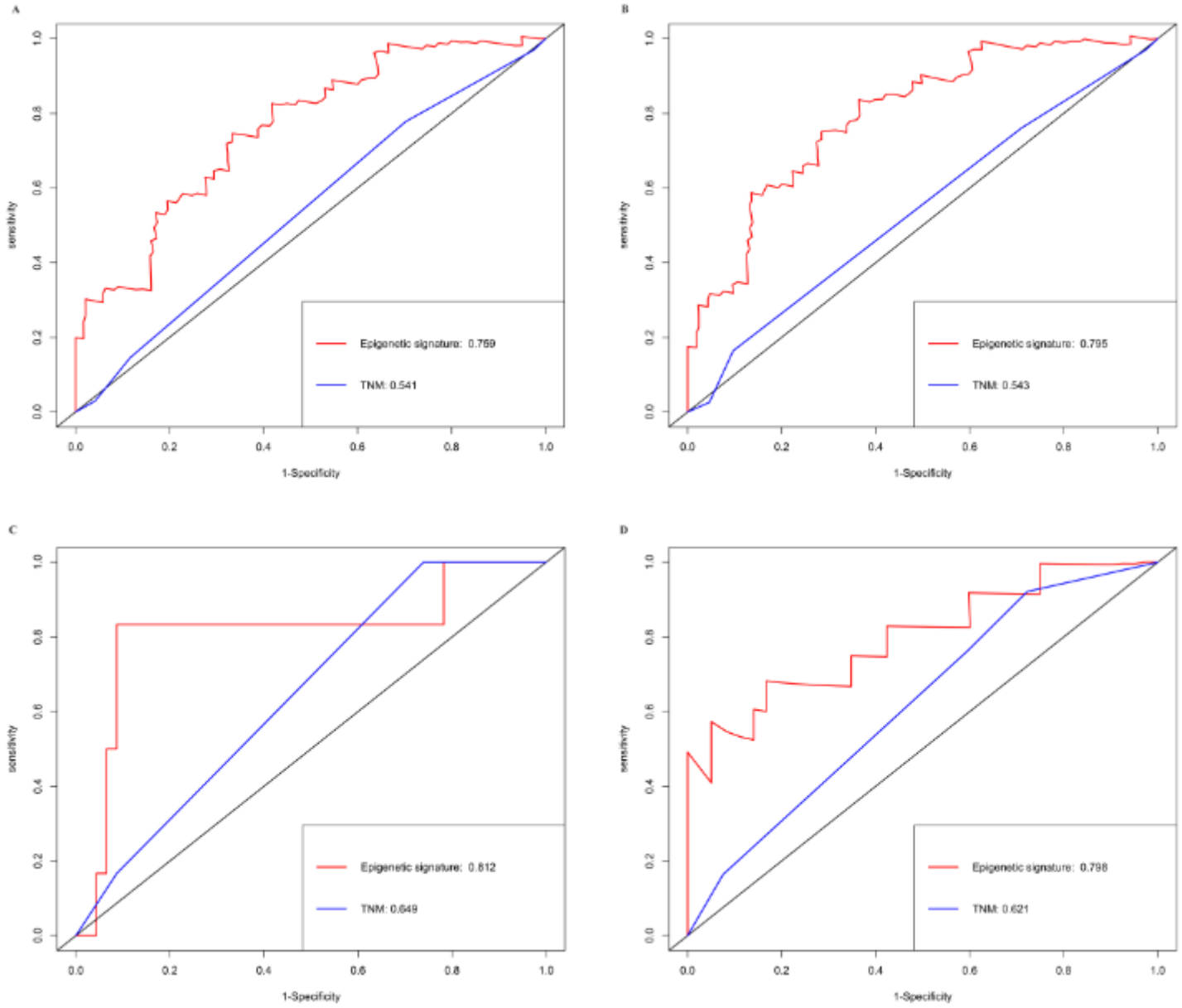

Figure 12

A and B: ROC curves compare the prognostic accuracy of the epigenetic signature with TNM staging in predicting 3-year (A) and 5-year (B) survival probability in the TCGA dataset. C and D: ROC curves compare the prognostic accuracy of the epigenetic signature with TNM staging in predicting 3-year (C) and 5-year (D) survival probability in the GEO dataset.

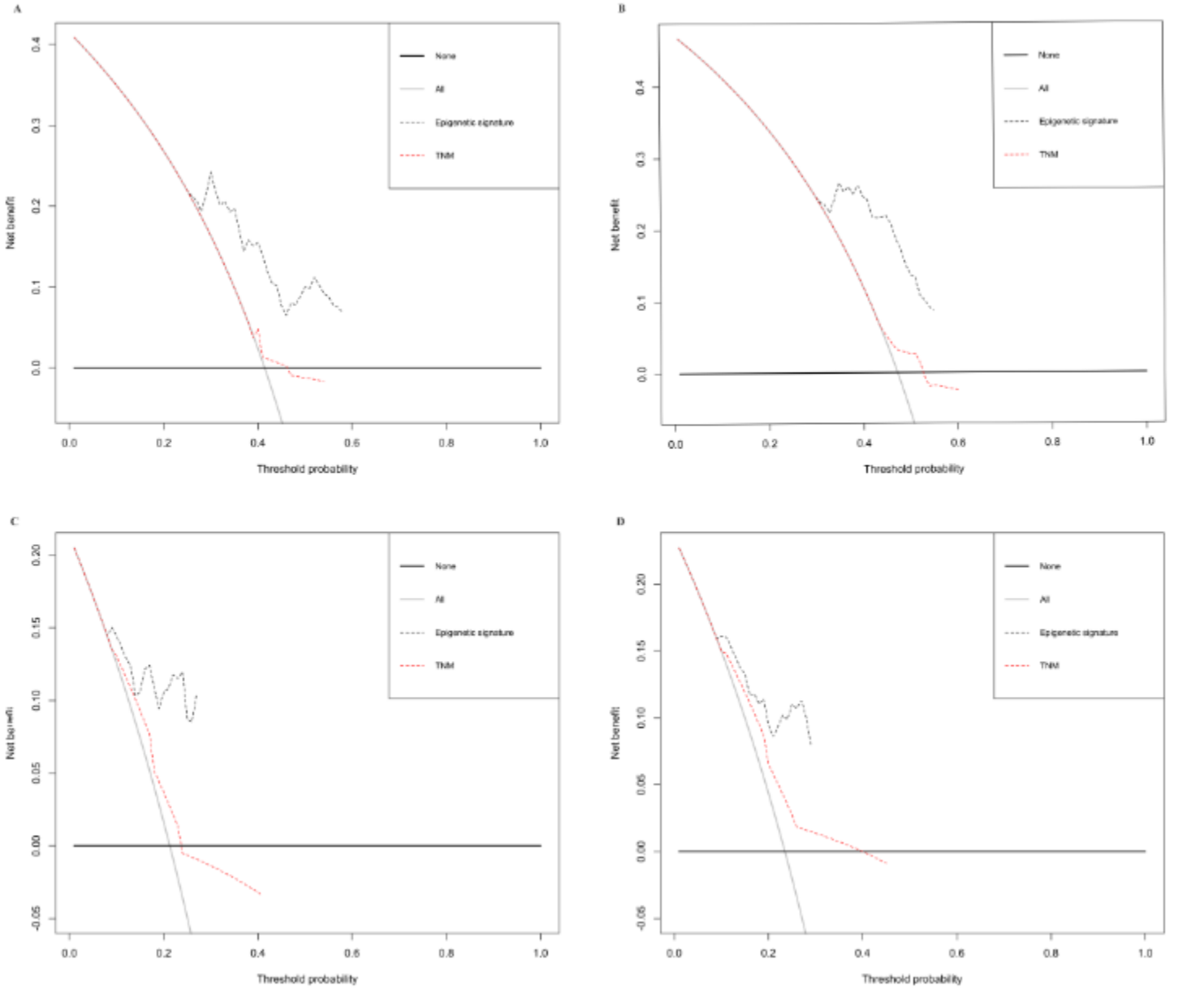

Figure 14

A and B: Decision curve analysis for the epigenetic signature and TNM staging in predicting 3-year (A) and 5-year (B) survival probability in the TCGA dataset. C and D: Decision curve analysis for the epigenetic signature and TNM staging in predicting 3-year (C) and 5-year (D) survival probability in the GEO dataset.

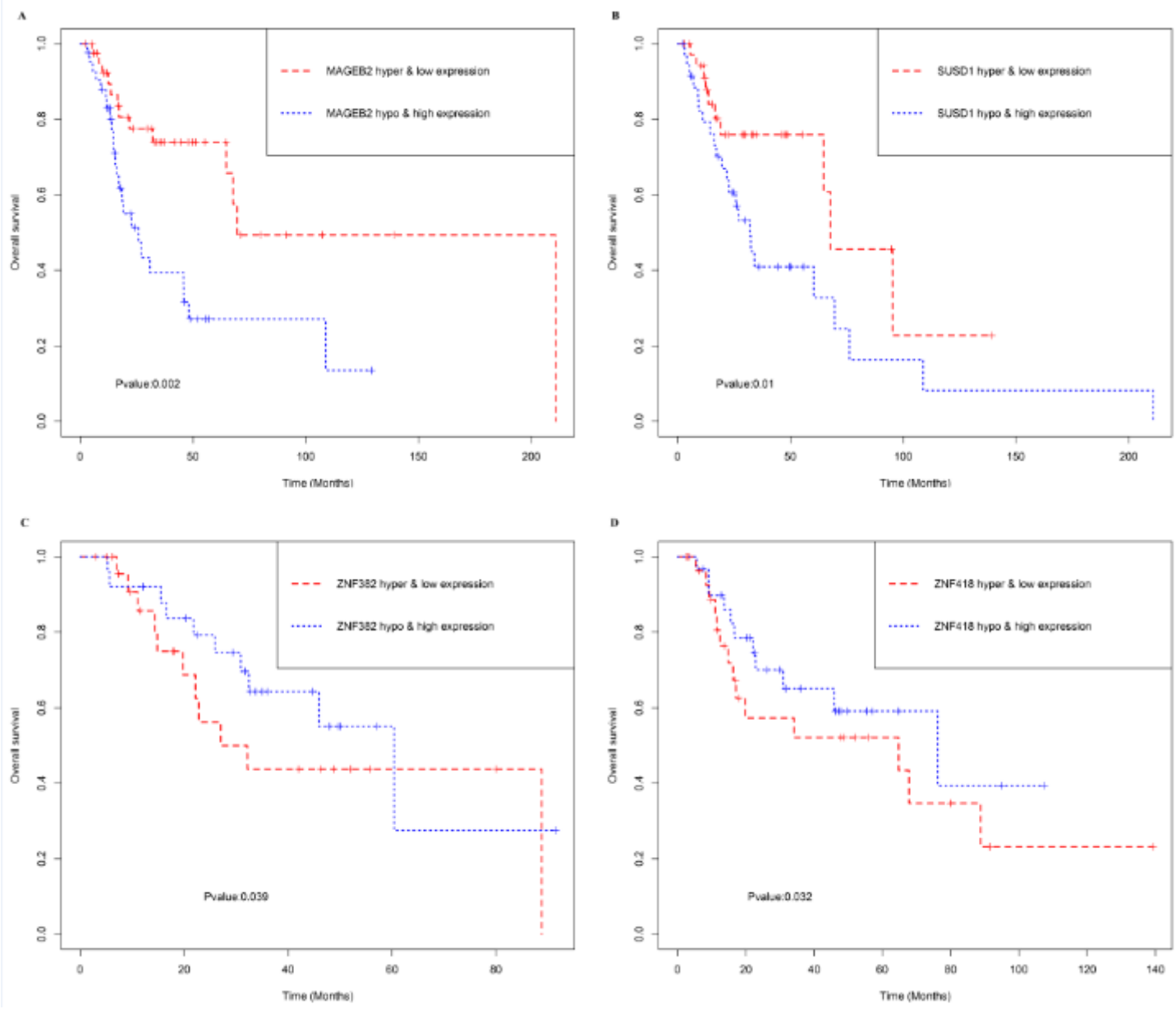

Figure 15

Kaplan–Meier survival curves for the joint survival analysis. A: The combination of gene MAGEB2 methylation and expression; B: The combination of gene SUSD1 methylation and expression; C: The combination of gene ZNF382 methylation and expression; D: The combination of gene ZNF418 methylation and expression.

Supplementary Files

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

- FigureS1.tif
- TableS1.docx

- Supplementarymaterial2.xls
- Supplementarymaterial1.xls
- FigureS2.tif
- Supplementarymaterial2.xls
- Supplementarymaterial1.xls
- FigureS4.tif
- FigureS5.tif
- FigureS2.tif
- FigureS4.tif
- TableS1.docx
- FigureS3.tif
- Supplementarymaterial3.xls
- FigureS1.tif
- FigureS3.tif
- FigureS5.tif
- Supplementarymaterial3.xls