Sixteen Cellular Senescence–associated DNA Methylation Signature Predicts Overall Survival in Patients with Head and Neck Squamous Cell Carcinoma

Ming Han YE[1], Xin Yi HUANG[2], Chun Jie LI[1,3], Qian Ju WU[4,5], Fei LIU[1]

Objective: *To construct a cellular senescence–related DNA methylation model to act as an independent prognosis predictor for patients with head and neck squamous cell carcinoma (HNSCC).*

Methods: *Methylome, transcriptome and clinical information for 499 HNSCC patients were received from The Cancer Genome Atlas (TCGA) as a training set. An extra independent methylation dataset of 54 patients with oral squamous cell carcinoma (OSCC) was downloaded from the NCBI Gene Expression Omnibus (GEO) database as the validation set. To assess the cellular senescence level of each sample, the senescence score (SS) of each patient was calculated using the transcriptome data via single-sample gene set enrichment analysis (ssGSEA). Least absolute shrinkage and selection operator (LASSO) Cox regression analyses were conducted to confirm Cytosine, phosphoric acid and Guanine (CpG) sites for the development of a cellular senescence–related DNA methylation signature.*

Results: *Based on the SS of each HNSCC patient in the TCGA cohort, the patients were divided into high- and low-SS subgroups. The high-SS group showed a better prognosis than the low-SS group. Moreover, 3,261 differentially methylated CpG sites (DMCs) were confirmed between the two groups. Among them, 16 DMCs were included to develop a 16-DNA methylation signature for evaluation of HNSCC prognosis using LASSO and multivariate Cox regression analysis.*

Conclusion: *A novel cellular senescence–related 16-DNA methylation signature was determined, which can be used as an independent index to evaluate the prognosis of HNSCC patients and select appropriate treatment strategies.*

Key words: *cellular senescence, diagnostic predictor, DNA methylation signature, head and neck squamous cell carcinoma, oral squamous cell carcinoma*

Chin J Dent Res 2023;26(4):235–248; doi: 10.3290/j.cjdr.b4784007

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common malignancy worldwide with a significant death rate and originates from the oral cavity, larynx and pharynx[1]. According to the 2020 Global Cancer Report, there more than 600,000 new HNSCC cases are diagnosed each year[2]. HNSCC is characterised as a heterogeneous malignancy with inter- and intra-tumour

1 The State Key Laboratory of Oral Diseases, National Clinical Research Centre for Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu, Sichuan Province, P.R. China.
2 School and Hospital of Stomatology, Cheeloo College of Medicine, Shandong University & Shandong Key Laboratory of Oral Tissue Regeneration & Shandong Engineering Laboratory for Dental Materials and Oral Tissue Regeneration, Jinan, P.R. China.
3 Department of Head and Neck Oncology, West China Hospital of Stomatology, Sichuan University, Chengdu, P.R. China.
4 Stomatological Hospital of Xiamen Medical College, Xiamen Key Laboratory of Stomatological Disease Diagnosis and Treatment, Xiamen, P.R. China.
5 Shanghai Ninth People's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, P.R. China.

Corresponding authors: Dr Qian Ju WU, Stomatological Hospital of Xiamen Medical College, Xiamen Key Laboratory of Stomatological Disease Diagnosis and Treatment, Xiamen 361008, P.R. China. Tel: 86-592-2678557. Email: qianjuwu@163.com

Dr Fei LIU, State Key Laboratory of Oral Diseases, National Clinical Research Centre for Oral Diseases, National Centre for Stomatology, West China School of Stomatology, Sichuan University, Chengdu 610041, P.R. China. Tel: 86-15680820580. Email: liufei.hxkq@scu.edu.cn

The present work was supported by the Department of Science and Technology of Sichuan Province (grant number: 2021JDRC0166) and the China Postdoctoral Science Foundation (grant number: 2020M683329).

differences[3-5]. Nowadays, although multiple therapies for HNSCC have developed significantly, the overall 5-year survival rate for HNSCC is still around 50%, which is far from satisfactory[6]. Accumulated evidence proved that early diagnosis of HNSCC leads to a higher 5-year survival rate and a better prognosis[6]. Hence, the discovery of trustworthy prognostic biomarkers is crucial to identify HNSCC at an early stage, offer patients effective therapy and improve their prognosis.

Cellular senescence is a permanent process of cell proliferation termination. It plays a pertinent role in tumour initiation growth, and metastasis, which is related to the activation of senescence-associated secretory phenotype (SASP), a key parameter generated by senescent cells[7-9]. Cellular senescence and SASP are a Jekyll and Hyde alteration, which means they facilitate suppression of the division of abnormal cells transforming into tumours while activating tumour cell growth and distant metastasis, particularly in tumours with higher SASP[9-10]. Studies have described how cellular senescence also plays a vital role in the resistance of HNSCC against radiotherapy and chemotherapy[11-13]. Investigating the particular mechanism underlying the regulation of cellular senescence on the HNSCC could lead to a new way of detecting HNSCC at an early stage and offering timely therapy. Thus, further research on cellular senescence in HNSCC is urgently needed.

DNA methylation is one of the most widely explored epigenetic modifications, exhibiting a tight correlation to tumours[14]. Recently, emerging evidence has supported the idea that DNA methylation regulates cellular senescence and expression level of cellular senescence-related genes (CSRGs) in HNSCC patients[12,15]. Accordingly, the exploration of DNA methylation and cellular senescence in HNSCC is potentially worthwhile.

Machine learning–based models show giant potential in evaluating tumour prognosis and improving the efficiency of immunotherapy in the treatment of tumours like HNSCC[16]. Recently, Wang et al[17] developed a cellular senescence–related prognosis model to predict the HNSCC survival rate. Nevertheless, to the best of the present authors' acknowledge, there is no HNSCC cellular senescence–related prognosis model constructed based on the methylome data.

The flowchart of the study workflow is shown in Fig 1. In the present study, in order to construct a cellular senescence–related DNA methylation model as a biomarker for the prediction of the prognosis of HNSCC patients, the methylome, transcriptome and clinical data for HNSCC patients were downloaded from the Cancer Genome Atlas (TCGA) to construct a prediction model of HNSCC prognosis. Sixteen DMCs were determined to develop a 16-DNA methylation signature for HNSCC prognosis evaluation by way of a least absolute shrinkage and selection operator (LASSO) and multivariate Cox regression analysis. This cellular senescence–related DNA methylation signature has been proven to be a significant independent prediction factor in the prognosis of HNSCC patients and the identification of a subgroup that is more suited for immunotherapy. The discoverymade in the present study may offer a better understanding of the role that cellular senescence plays in the prognosis of HNSCC and provide new strategies to identify subgroups and offer accurate therapy for HNSCC.

Materials and methods

Data collection and pre-processing

The data of HNSCC patients included in this study were obtained from UNSC Xena (https://xena.ucsc.edu/). The online database provided the transcriptome (RNA-seq, HTSeq-Counts/FPKMs) and methylome (Illumina Human Methylation 450, HM450) data containing 566 samples along with corresponding phenotype information. A total of 499 HNSCC patients and 44 normal controls with complete corresponding survival data were selected as the training set. The expression value (FPKM) was then normalised by quantile normalisation. An extra independent methylation dataset including 54 OSCC samples was downloaded from the NCBI Gene Expression Omnibus (GEO) database (access no. GSE75537)[18]. The GEO dataset was downloaded for external verification as a validation set.

Inference of senescence score (SS) and survival analysis investigating the relationship between SS and HNSCC prognosis

A total of 279 cellular senescence-related genes (CSRGs) were collected from the CellAge database including 28 oncogene-induced, 232 replicative and 34 stress-induced genes[19]. Five of these were both oncogene-induced and replicative, and ten were both stress-induced and replicative. To further understand the role played by cellular senescence in HNSCC and to investigate the methylation sites associated with cellular senescence and the relationship between cellular senescence and patient prognosis, the SS model was introduced. Firstly, a univariate Cox survival analysis was performed to confirm the CSRGs of which the expression level showed positive

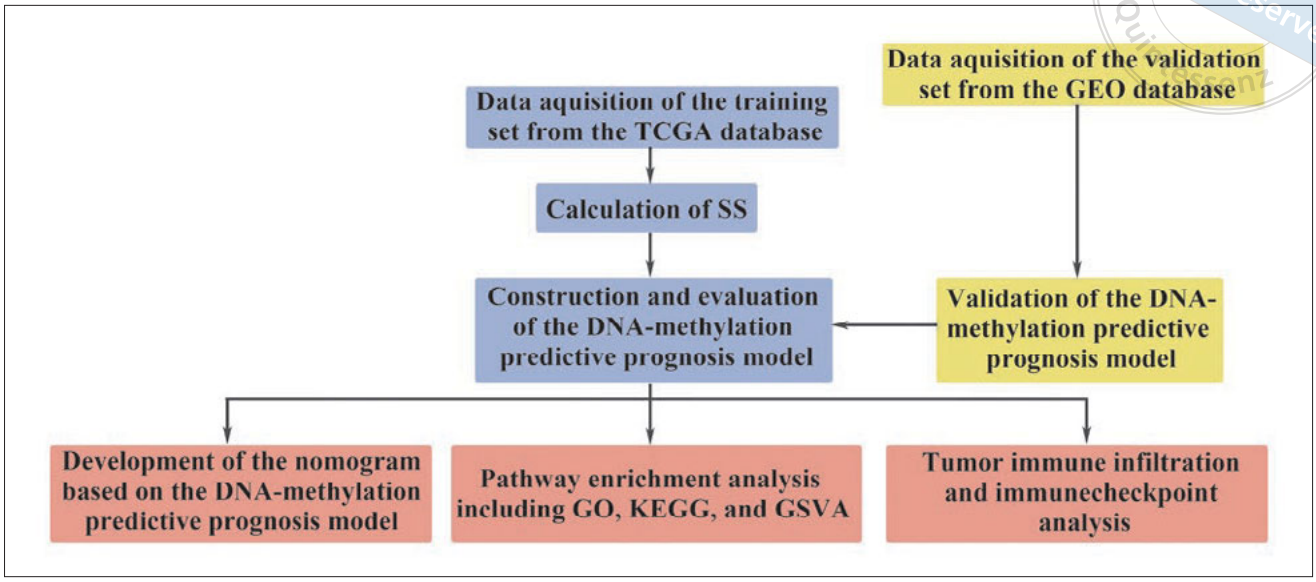

Fig 1 Flowchart for the present study.

and negative correlations with HNSCC patient prognosis, which were then defined as the senescence-positive and negative gene sets, respectively. The SSs were then inferred according to the method described in two previous reports[20-21]. Enriched SSs based on the two gene sets above were computed by applying ssGSEA in the Gene Set Variation Analysis (GSVA) R package[22], and the enriched score of senescence-positive components minus that of senescence-negative components was calculated and defined as the SS for each HNSCC individual. To verify the relationship between SS and HNSCC prognosis, patients in the training set were divided into high- and low-SS groups, which were confirmed by the median SS of the 499 TCGA HNSCC patients. In this study, a higher SS represented higher cellular senescence activity.

The survival status, survival time and phenotype data including the age, gender, sex and race of TCGA HNSCC patients were extracted to determine whether cellular senescence is related to the prognosis of patients with HNSCC via univariate and multivariate Cox proportional hazards analyses using the survival R package. Kaplan-Meier (KM) survival curves were drawn using the "survfit" function of the survminer R package to show the difference in the prognosis of HNSCC patients between the high- and low-SS groups.

Analysis of DMCs related to tumorigenesis and HNSCC prognosis

DMCs between tumour and normal samples in the training set were initially extracted using the limma R package, then the DMCs between the high- and low-SS subgroups were confirmed. The DMCs between HNSCC samples and normal samples were determined with a *P* value < 0.01 and |logFC| > 1.2. The DMCs between high- and low-SS groups were determined with a *P* value < 0.01 and |logFC| > 1.1.

Construction and validation of the predictive model of HNSCC prognosis

To develop a prognosis signature, the survival time, survival status and methylome data of DMCs included in the previous section were integrated to conduct a LASSO Cox regression analysis using the glmnet R package. By way of 10-fold cross-validation, the appropriate coefficient value was calculated to create the following predictive model of HNSCC prognosis:

$$\text{Risk score} = \sum_{i=1}^{n} coef(i) * expr(i).$$

coef was the coefficient value of each corresponding DMC, and expr was the expression level of each corresponding DMC. To validate the prediction efficacy of the predictive model, the validation set from the GEO database was selected. Specifically, the risk score for each patient in the validation set was calculated and divided into high- and low-risk score groups according to the median risk score. In both the training set and the validation set, KM and ROC analyses were performed to examine the prediction ability. ROC analysis was carried out based on the pROC R package to obtain the area under the ROC curve (AUC) at 1, 3 and 5 years.

Development of a nomogram

Using the "rms" R package, the survival time, survival status and eight clinical features were integrated to develop a nomogram and calibration curves. The nomogram is available when evaluating the prognosis significance of each variable. Total points can be calculated by adding up all the values of the corresponding features in this nomogram to evaluate patients' prognoses. ROC curves were drawn to confirm the prediction ability of this nomogram to act as an independent predictor. The method of ROC analysis was described in the previous section.

Gene function enrichment analysis

To explore the potential biomolecular mechanisms underlying the regulation of the CSRGs on the prognosis of HNSCC patients, GO, KEGG and GSVA analyses were performed. First, differential expression analysis was conducted to identify differentially expressed genes (DEGs) between the high- and low-risk groups. Genes with $P < 0.01$ and $|\log FC| > 1.5$ met the criteria. To extract essential pathways that were closely correlated to the DEGs, GO and KEGG gene function enrichment analysis were performed in clusterProfiler. GO analysis is a widely popular method that annotates genes and downstream products with three groups: molecular function (MF), biological pathways (BP) and cellular components (CCs)[23]. GO annotation information was extracted from org.Hs.eg.db. KEGG is a powerful and comprehensive database developed for the systematic investigation of gene function pathways[24-25] based on the newest KEGG pathway annotation from KEGG rest API (https://www.kegg.jp/kegg/rest/keggapi.html). To evaluate the influence of the expression level of each DEG, GSVA analysis[22] was performed based on GO and KEGG terms downloaded from the Molecular Signatures Database separately[26].

Assessment of tumour immune infiltration and immune checkpoints in HNSCC

To clarify the tumour immune purity in HNSCC, the tumour microenvironment in each TCGA HNSCC sample was then explored. ESTIMATE[27] and TIMER[28] methods were used to clarify tumour immune purity in detail via the IOBR R package[29].

Considering the potential prospect of immune checkpoints in immunotherapy for HNSCC, a list of known immune checkpoints were made according to the previous literature[30], and their expression level in high- and low-risk groups was detected to identify possible therapeutic targets.

Statistical analysis

In the current research, the online data analysis platform Sangerbox 3.0 (http://vip.sangerbox.com/) was used to conduct the aforementioned statistical analysis and result visualisation. Sangerbox 3.0 is a comprehensive, interaction-friendly website platform for bioinformatics analysis that makes it possible to perform a series of customized bioinformatics analyses and visualisation mapping based on the R packages mentioned above, and corresponding results can be downloaded from the platform[31].

Results

SSs of HNSCC patients were associated with their prognosis

To investigate the role of cellular senescence in the genesis of HNSCC, the CSRGs were retrieved from the CellAge database[19]. Next, 11 and 9 genes (Table 1) were distinguished showing better and worse overall survival in the TCGA cohort, respectively, between expression level and overall survival ($P<0.01$). The two gene sets were defined as senescence-positive and senescence-negative. Based on the expression levels of the two gene sets using ssGSEA, the SSs of HNSCC patients were calculated. SS represents the cellular senescence activity of different individuals to assess whether there is a relationship between cellular senescence and HNSCC patients, and was used in the subsequent section for screening of differentially expressed DNA methylation sites. The SSs ranged between −0.61 and 1.00 and the median was 0.16. Next, the HNSCC patients were separated into high- and low-SS groups by the median SS and a KM survival analysis was conducted. Using a log-rank test, the findings indicated that low-SS HNSCC patients had poor overall survival ($P < 0.01$; Fig 2a). After adjusting for the effects of variables, such as age, sex, race and clinical stage, a multivariate Cox model analysis was performed and the relationship between SSs and survival of patients with HNSCC was verified (Fig 2b). These results revealed that SSs of the TCGA HNSCC samples were associated with their prognosis.

Table 1 CSRGs related to HNSCC survival confirmed by the training set.

Genes (ensemble ID)		Hazard ratio (HR) (95% CI for HR)	*P* value
Senescence-positive gene sets	MAP2K7 (ENSG00000076984.16)	0.895 (0.843–0.951)	0.0003729
	MXD4 (ENSG00000123933.13)	0.937 (0.905–0.971)	0.0006764
	MAP3K6 (ENSG00000142733.13)	0.967 (0.948–0.986)	0.0009821
	CDKN2A (ENSG00000147889.15)	0.982 (0.971–0.992)	0.0010847
	MAP4K1 (ENSG00000104814.11)	0.877 (0.807–0.952)	0.0017620
	BTG3 (ENSG00000154640.13)	0.974 (0.956–0.990)	0.0021045
	LGALS3 (ENSG00000131981.14)	0.995 (0.991–0.998)	0.0021653
	DUSP16 (ENSG00000111266.7)	0.943 (0.907–0.981)	0.0037633
	BLK (ENSG00000136573.11)	0.600 (0.420–0.857)	0.0038601
	MOB3A (ENSG00000172081.12)	0.976 (0.959–0.993)	0.0068108
	EHF (ENSG00000135373.11)	0.991 (0.984–0.998)	0.0091848
Senescence-negative gene sets	PCGF2 (ENSG00000277258.3)	1.055 (1.025–1.086)	0.0003175
	FXR1 (ENSG00000114416.16)	1.035 (1.015–1.055)	0.0004351
	MAP2K1 (ENSG00000169032.8)	1.038 (1.015–1.061)	0.0011761
	SLC13A3 (ENSG00000158296.12)	1.484 (1.147–1.921)	0.0026281
	PSMD14 (ENSG00000115233.10)	1.050 (1.016–1.085)	0.0037126
	SMURF2 (ENSG00000108854.14)	1.130 (1.040–1.228)	0.0037972
	AURKA (ENSG00000087586.16)	1.031 (1.009–1.053)	0.0062148
	ERRFI1 (ENSG00000116285.11)	1.013 (1.003–1.022)	0.0075066
	TERF2 (ENSG00000132604.9)	1.123 (1.029–1.226)	0.0093966

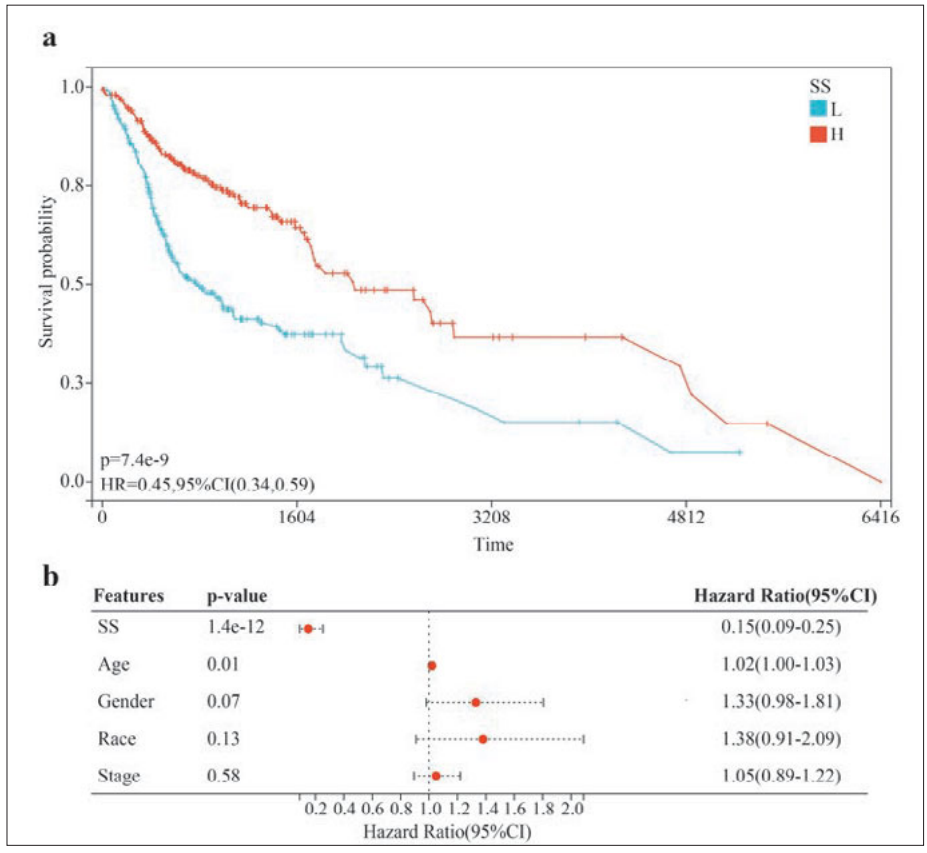

Fig 2 Relationships between SS and HNSCC overall survival. **(a)** KM curves of HNSCC with high- and low-SS subgroups in the training set. **(b)** Forest plot of the relationship between SS and HNSCC survival based on a multivariate Cox analysis.

The predictive prognosis model based on 16 cellular senescence–related DMCs was developed based on HNSCC overall survival

In the TCGA cohort, 12,046 DMCs were initially uncovered between HNSCC patients and normal samples ($P < 0.01$ and $|logFC| > 1.2$). Ulteriorly, 3,261 DMCs were identified between high- and low-SS subgroups ($P < 0.01$ and $|logFC| > 1.1$).

Next, the prognosis value of the 3,261 DMCs in the training set was evaluated using the LASSO Cox regression model. Through ten-fold cross-validation, the optimal lambda value was selected and then the range of DMCs was condensed in order to find a DNA methylation signature for the prediction of overall survival in HNSCC. As a result, a total of 16 CpG sites were included to construct the predictive prognosis model (Figs 3a and b). Using each CpG site's coefficient value and methylation status, a risk score model was developed as follows: RiskScore = (0.4921997969551*cg01995815) + (0.24140263854671*cg02409878) + (-0.074692525763 4289*cg03424345) + (-0.185126888295284*cg06459070) + (-0.084858590961658*cg06879746) + (0.024014297893 4278*cg06903569) + (-0.4051941 67436796*cg07768107) + (-0.061251172529722*cg08106706) + (-0.049130588721 8784*cg12298745) + (0.430730131248078*cg12389461) + (0.153090807571048*cg13757826) + (-0.0119448628610 951*cg14170959) + (0.301928975546258*cg16123269) + (0.243110636620704*cg19239041) + (-0.07440882137424 15*cg20711812) + (-0.978573233155824*cg21230425). To elucidate the efficacy of the influence of each CpG site included in the predictive prognosis model in the overall survival of HNSCC, multivariate and univariate Cox analyses were conducted. The multivariate Cox model was then used to compute the coefficient values of these 16 CpG sites (Fig 3c). Additionally, using univariate Cox regression, the relationship between each CpG site and HNSCC survival was assessed (Fig 3d). Based on the results of the multivariate and univariate Cox models, the methylation of the 16 CpG sites was strongly linked with HNSCC survival. A Pearson correlation analysis was also performed to investigate whether risk scores and SSs of the HNSCC patients in the training set were significantly related. A significant negative correlation was observed between the risk scores calculated by the 16-DNA methylation predictive prognosis model and SSs in the TCGA dataset ($P = 4.2e-37$; Fig 3e).

The potential impact of the 16 CpG sites of the DNA methylation signature on the 20 CSRGs associated with HNSCC survival was also investigated. A Pearson correlation analysis was conducted to investigate the relationship between the 16 CpG sites and 20 CSRGs impacting HNSCC prognosis. The results revealed that all the DMCs selected were significantly correlated with several CSRGs ($P < 0.01$; Fig 3f). Among them, the methylation level of cg21230425 significantly correlated with the maximum number of CSRGs amounted to 16, whereas that of cg14170959 correlated with the minimum of the CSRGs amounted to 2. These results imply that the 16 DMCs probably regulate CSRG transcription either directly or mediately.

The 16-DMC predictive prognosis model is an independent predictor for the prognosis of HNSCC patients

Each HNSCC patient's risk score was calculated and then the patients were separated into high- and low-risk subgroups according to the median score to analyse the survival prognosis model. Following this, a KM survival analysis was performed, and found that HNSCC patients in the high-risk subgroup had a significantly worse probability of survival ($P = 3.4e-12$; Fig 4a), suggesting that the risk score model could be a reliable prognostic predictor. As shown in Fig 4b, compared to the low-risk subgroup, more deaths occurred in the high-risk subgroup. The heatmap visualised the methylation value of the 16 DMCs in the risk score signature. To assess the prediction ability of the predictive prognosis model, an ROC analysis was conducted. For patients with HNSCC from the training set, this model was quite effective at predicting survival, with AUC of 0.69, 0.73 and 0.75 for predicting 1-, 3- and 5-year survival rate, respectively (Fig 4c). In summary, using the TCGA cohort, the 16-DMC predictive prognosis model was confirmed to be an independent parameter for the prognosis of HNSCC patients.

16-DMC predictive prognosis model is validated in the OSCC GEO dataset

As described above, a 16-DNA methylation signature with an authentic effect in predicting the prognosis of HNSCC patients was constructed based on the TCGA database. Since OSCC is the main disease type of HNSCC, methylation and survival data from the OSCC dataset GSE75537 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) to further verify the risk model. As expected, OSCC patients in the GEO dataset with higher risk scores had a worse prognosis ($P < 0.01$; Fig 5a). Based on the developed prognostic model, heatmap visualisation of different DMCs for patients and survival distribution in the GEO dataset also showed a similar pattern as in the TCGA dataset (Fig 5b), with an

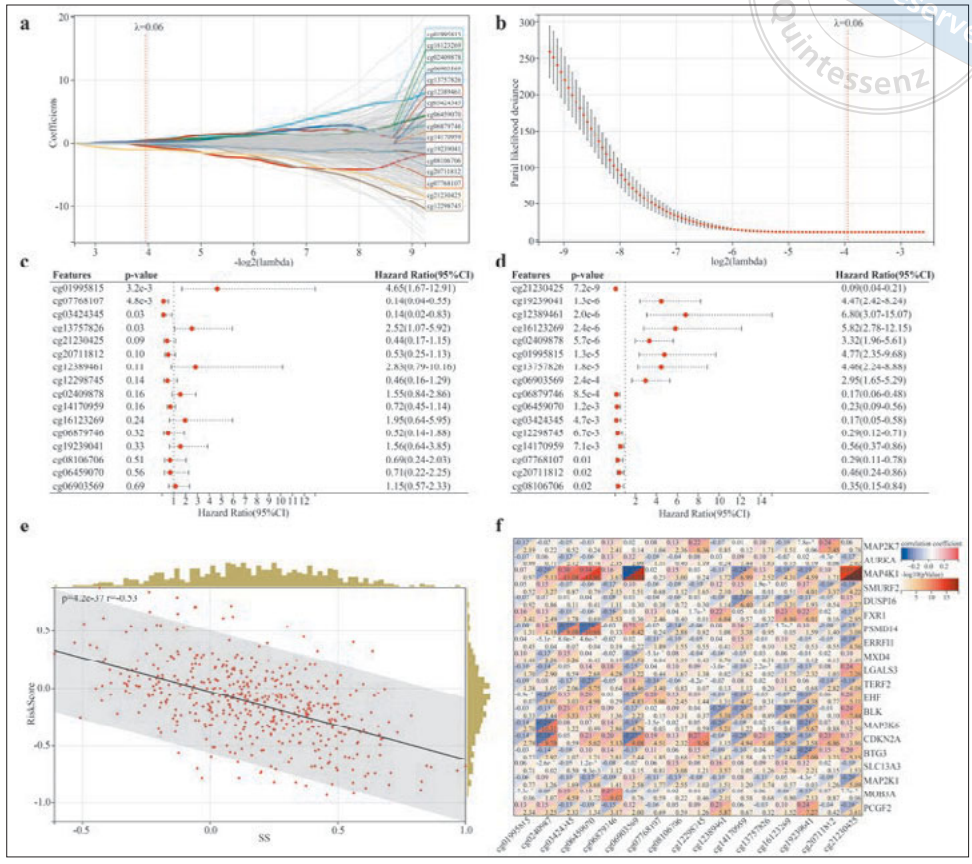

Fig 3 Constructing a prognostic model. **(a)** LASSO regression coefficients. **(b)** Lambda value was selected to identify genes and develop a prognosis model. **(c)** A multivariate Cox analysis showed the correlation between methylation values. **(d)** A univariate Cox analysis showed the correlation between the methylation value of the 16 DMCs and TCGA HNSCC patients' survival. **(e)** A Pearson correlation analysis showed the relevance between SS and risk score in the training set. **(f)** Relevance between methylation value of the 16 DMCs selected to construct the prognosis signature and expression level of CSRGs impacting HNSCC survival.

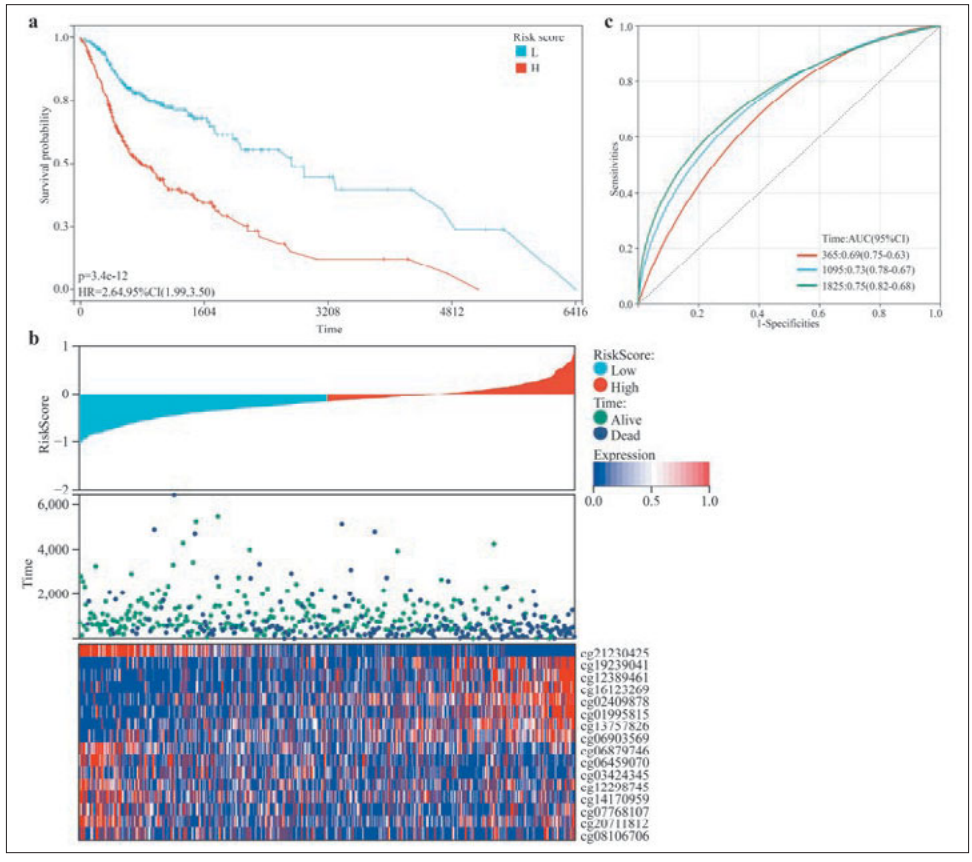

Fig 4 Development of cellular senescence–related DNA methylation prognosis model based on methylation level of 16 DMCs in the training set. **(a)** KM curves for high- and low-risk subgroups in the training set. **(b)** Risk survival status plot and heatmap of the methylation level of the 16 DMCs model in TCGA HNSCC patients with high- and low-risk scores. **(c)** AUCs for the prediction of 1-, 3- and 5-year overall survival of HNSCC.

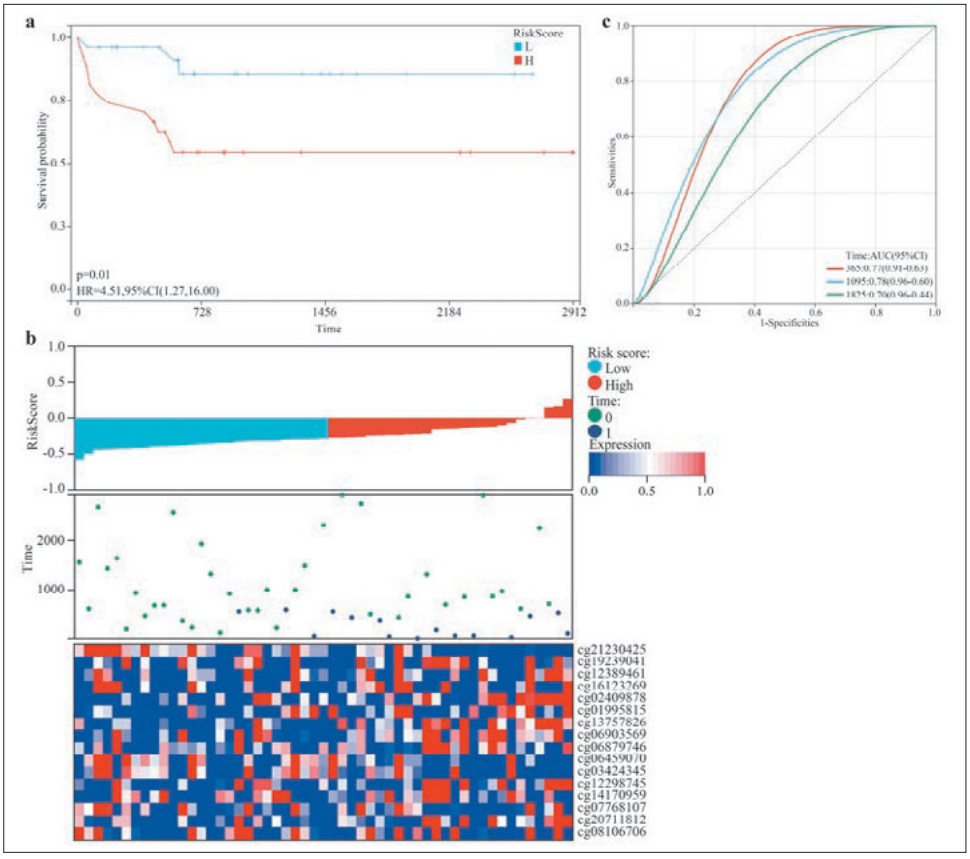

Fig 5 Validation of cellular senescence–related DNA methylation prognosis model based on methylation level of 16 DMCs in the validation set. **(a)** KM curves for high- and low-risk subgroups in the GEO cohort. **(b)** Risk survival status plot and heatmap of the methylation level of 16 DMCs model in GEO OSCC patients with high- and low-risk scores. **(c)** AUCs for the prediction of 1-, 3- and 5-year overall survival of OSCC.

AUC of 0.77 at 1 year, 0.78 at 3 years and 0.70 at 5 years, respectively (Fig 5c). Thus, the validation results indicate that the model is sufficiently stable to be applicable to a broader HNSCC population.

Establishment of a nomogram based on the model with predictive efficacy in HNSCC prognosis

To thoroughly predict HNSCC patients' survival based on both risk score and clinical factors, a nomogram was developed by age, sex, grade, clinical stage, clinical T stage, clinical M stage, clinical N stage and risk score (Fig 6a). When the total points were 53.06, the predicted survival probability was more than 0.90 at 1 year, more than 0.80 at 3 years and more than 0.75 at 5 years. Subsequently, it was proved that the observed survival probability and predicted survival probability were almost in congruity on the 3- and 5-year calibration curves (Fig 6b). The result proved that this nomogram is highly accurate in predicting the survival probability of HNSCC patients. ROC curves were then depicted (Fig 6c). The 1-, 3- and 5- year AUCs were 0.69, 0.75 and 0.77, respectively. This validated the excellent survival rate prediction ability of the nomogram.

Discovery of differential pathways related to immune function between the high- and low-risk subgroups of HNSCC

Subsequently, differential expression analysis using the limma package was conducted; thus, 4,673 DEGs were uncovered in the two subgroups ($P < 0.01$, $|\log FC| > 1.5$). GO and KEGG enrichment analyses were performed to crystallise the function and mechanisms of the preceding DEGs. The enrichment results were visualised in bubble plots and the findings of the GO analysis included three different function types: BP, CC and MF (Figs 7a to d).

In the BP sections, the previous DEGs exhibited enrichment in "response to stress", "immune system process" and "immune response". For CC, the most significantly enriched terms contained "cytosol", "endomembrane system" and "vesicle". In MF terms, the DEGs were significantly enriched in "catalytic activity" and "enzyme binding" along with "identical protein binding". Additionally, the bubble plot for the KEGG analysis revealed that these genes were enriched in the top 10 pathways, among which "cell adhesion molecules (CAMs) signalling pathway", "Th17 cell differentiation signalling pathway", "hematopoietic cell lineage

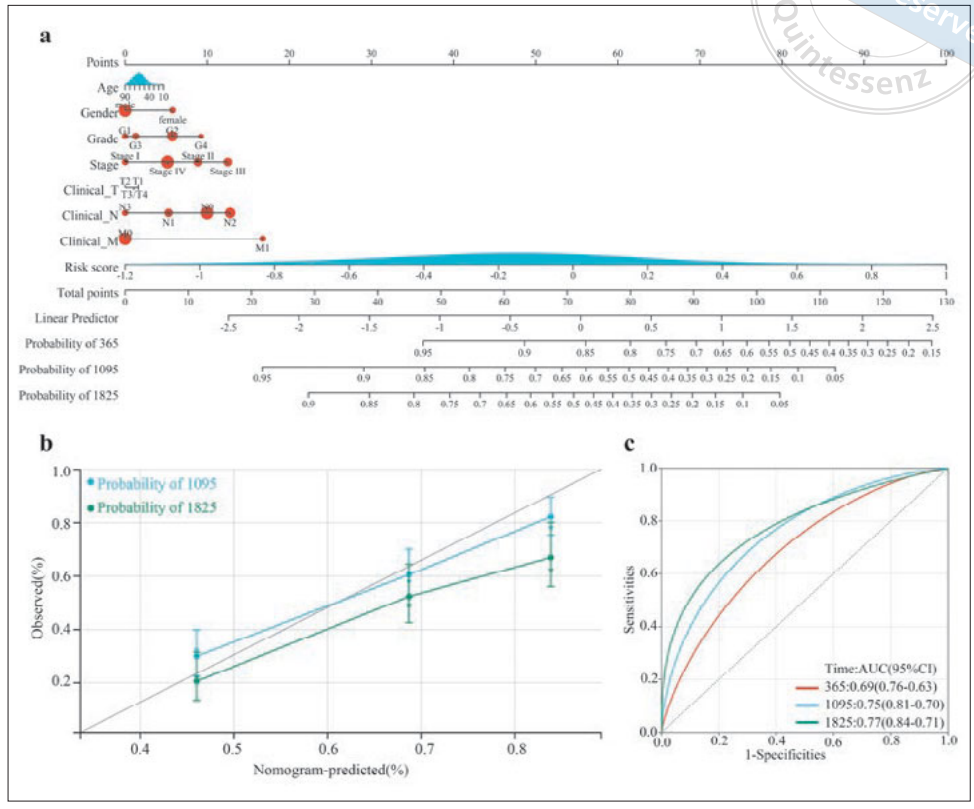

Fig 6 Development of a nomogram for survival prediction. **(a)** Nomogram with sufficient predictive efficacy. **(b)** 3- and 5- year calibration curves of the nomogram. **(c)** ROC curves of the nomogram.

signalling pathway" and "primary immunodeficiency signalling pathway" were closely related to immunity function.

Considering the expression level of DEGs may influence the enrichment results, GSVA analysis was performed to further clarify the role these DEGs played in the HNSCC patients. The heatmap contained the top 20 KEGG signalling pathways enriched by GSVA (Fig 7e). According to the analysis, signalling pathways such as "Glycosphingolipid biosynthesis-ganglio series", "Glycosaminoglycan biosynthesis-chondroitin sulphate", "Glycosaminoglycan biosynthesis-keratan sulphate" and "Glycosaminoglycan biosynthesis-heparan sulphate" were significantly active in HNSCC patients in the high-risk subgroup, suggesting the genesis and development of HNSCC were accompanied by an abnormal change in glycan biosynthesis and metabolism. Meanwhile, pathways related to the immune system, such as the "B cell receptor signalling pathway" and "intestinal immune network for IgA production and T cell receptor signalling pathway" were significantly suppressed. All the pathway enrichment results indicated the possibility that abnormal immune system function plays an important role in the prognosis of HNSCC patients.

Immune infiltration CSRG transcription/DNA methylation is correlated with HNSCC

Given the crosstalk between cellular senescence and tumour immune infiltration, the authors then investigated whether the cellular senescence–related 16-DMCs model played a role in immune cell infiltration of HNSCC patients. ESTIMATE and TIMER algorithms were used to clarify the tumour purity and immune infiltration level in the HNSCC tumour microenvironment. The immune infiltration level between the high- and low-risk subgroups was compared and violin plots were drawn to elucidate whether the discrepancies found in ESTIMATE and TIMER immune scores existed between the two groups (Figs 8a and 8b). As shown in Fig 8a, the immune score and ESTIMATE score were significantly lower in the high-risk group compared to the low-risk group ($P < 0.01$). Meanwhile, there was no significant difference in the stromal score found in the two groups ($P = 0.77$). These findings revealed that immune cell infiltration may decrease in the tumour microenvironment in HNSCC patients. These results were in accordance with the TIMER analysis. As shown in Fig 8b, five types of immune cells except macrophages exhibited significantly lower infiltration in the high-risk subgroup ($P <$ 0.01). These indications supported the possibility that

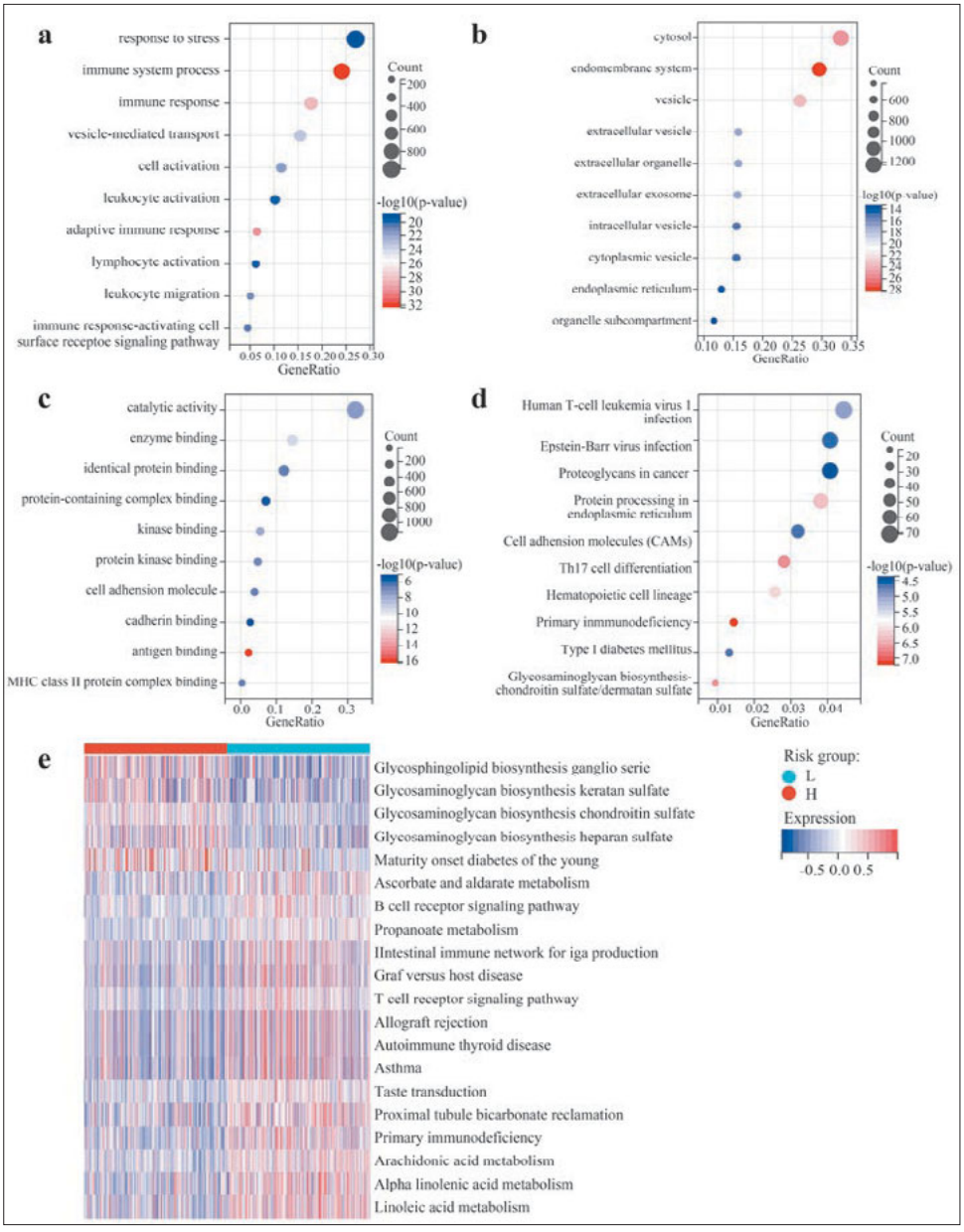

Fig 7 GO and KEGG analysis of the DEGs. **(a)** Results of GO-BP. **(b)** Results of GO-CC. **(c)** Results of GO-MF. **(d)** Results of KEGG. **(e)** Results of GSVA.

immune infiltration influences the prognosis of HNSCC patients.

To further confirm this, a Pearson correlation analysis was performed to evaluate the relationship between risk score and immune infiltration level (Figs 8c and 8d). With the exception of the stromal score calculated using the ESTIMATE algorithm and the macrophage infiltration score calculated using the TIMER algorithm, the results demonstrated a significantly negative relationship between risk score and all the immune scores ($P < 0.01$). Thus, the present research proved that the 16-DMCs prognosis signature played an indispensable role in decreasing the infiltration level of immune cells in the tumour microenvironment in HNSCC, which shed a light on immune therapy for HNSCC.

The variation in immune checkpoint expression between the two subgroups was also investigated in light of the significance of checkpoint-based immunotherapies. A list of immune checkpoints was retrieved from previous studies to compare their expression levels in the two groups (Fig 8e)[30]. Of the 11 immune checkpoints, 9 checkpoints (BTLA, HAVCR2, LAG3, LILRB2, PDCD1, SIGLEC7, SIRPA, TIGHT and VSIR) were confirmed to express significantly lower levels in the high-risk subgroup, which may provide a new immunotherapy strategy that involves choosing effective targeted medicine.

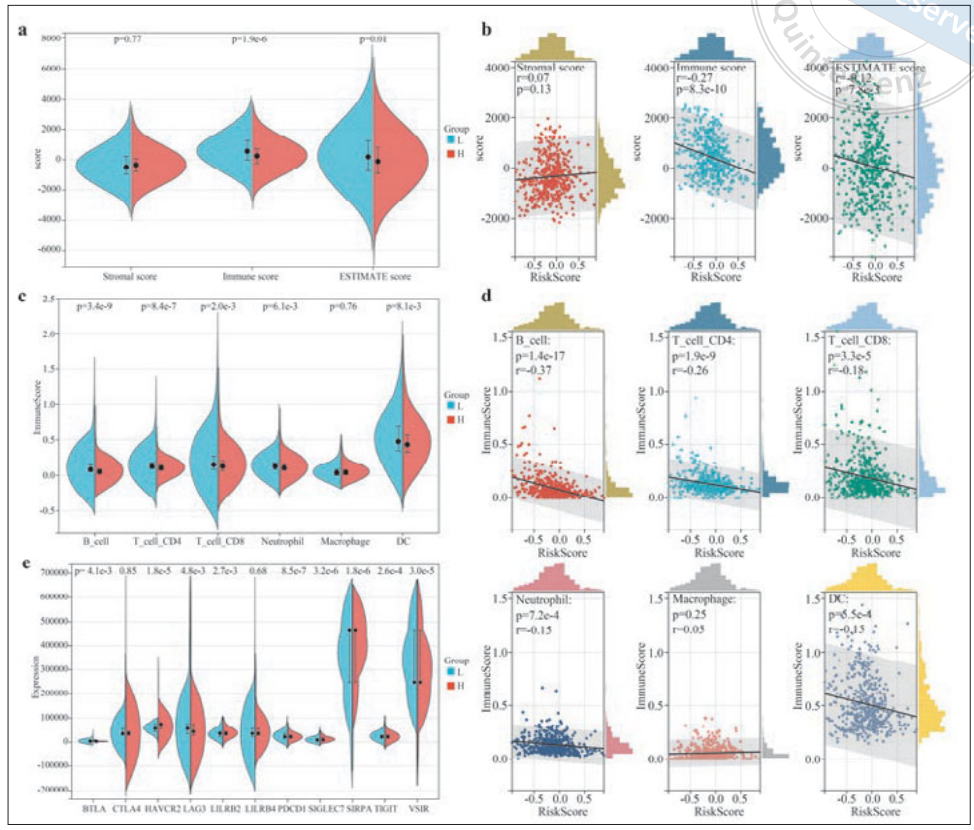

Fig 8 Immune infiltration and immune checkpoint analysis. **(a)** Immune infiltration analysis result by ESTIMATE. **(b)** Immune infiltration analysis result by TIMER. **(c)** A Pearson correlation analysis showed the relevance between risk score and ESTIMATE immune infiltration level in TCGA HNSCC patients. **(d)** A Pearson correlation analysis showed the relevance between risk score and TIMER in HNSCC patients. **(e)** Expression level of immune checkpoints in low- and high-risk subgroups.

Discussion

Since the discovery of cellular senescence[19], the phenomenon has been widely investigated and has been proven to be involved in oncogene activation and tumour suppression[32-33]. The most recent research has demonstrated that senescent cancer cells are heterogeneous due to various variables, manifesting both pro- and anti-tumour effects, suggesting that cellular senescence may lead to novel treatment strategies[34-35]. DNA methylation, a well-studied form of epigenetic modification involving the attachment of a methyl group onto the C5 position of the cytosine to synthesize 5-methylcytosine, influences the transcription of corresponding genes (like lysine methyltransferases [KMTs]) in various tumour types, which alter cellular senescence in tumour tissue and may serve as biomarkers in HNSCC prognosis[36-38]. In the present study, the 16-DNA methylation HNSCC predictive prognosis model was first developed based on the SS and was further validated. Further analysis showed that immune system function change and tumour immune microenvironment alteration occur in HNSCC, which is related to HNSCC prognosis and cellular senescence. The results indicate that the 16 DMCs related to cellular senescence may regulate the local lesion by affecting immune function. The present results could be profound in the HNSCC therapy area.

First, the methylome and transcriptome data were collected from the TCGA database for HNSCC patients. Based on the expression of 20 CSRGs and HNSCC survival, a unique SS was first derived for each HNSCC patient. The method used to infer sample scores followed that employed in previous articles on the relationship between cancer and another biological process[39,40]. Among the CSRGs involved, several genes have been confirmed that participate in cellular senescence and tumorigenesis. For instance, FXR1 is considered a promoter for enhancing cellular senescence[41], and its regulation activity is conducive to tumour growth in different cancer tissues including HNSCC[42-44]; CDKN2A is commonly altered in human tumours, which influences tumour-immune microenvironments[45] and plays an important role in in vivo aging via the p19(Arf)-p53 pathway[46]. The KM survival analysis revealed that HNSCC patients with higher SS had a poorer prognosis ($P < 0.01$), indicating a strong correlation between SS and HNSCC survival.

To create a DNA methylation predictive prognosis model for the diagnosis and prediction of HNSCC, 3,261 DMCs were next identified between the high- and low-SS groups in the TCGA. Ultimately, 16 DMCs were included via the LASSO Cox regression method. Based

on this model, the risk score for each HNSCC patient was calculated. The Pearson correlation analysis showed a significant negative relationship between SS and risk score. ROC curves revealed that the DNA methylation model was able to predict HNSCC prognosis reliably. Its prediction accuracy and stability were validated via a further investigation in the dataset GSE75537[18]. Furthermore, the DMCs used in model construction were proven to be closely related to the expression of the aforementioned CSRGs, thus indicating the possible interaction with CSRGs in cancer tissue. To include the effect of clinical variables and increase the accuracy in prognosis prediction, a nomogram was constructed based on the 16-DNA methylation risk model. ROC curves and calibration curves both confirmed the accurate evaluation ability in HNSCC patients.

To further clarify the concrete mechanism in HNSCC, GO[23] and KEGG[24,25] analyses were conducted to enrich signaling pathways using DEGs in high- and low-risk subgroups. The results proved that immune system function change may be involved in tumorigenesis and disease development, like "immune system process", and "immune response" in GO terms. Through GO enrichment, several terms were identified, some of which have been reported in the previous literature to regulate cellular senescence in tumours. For instance, extracellular vesicle (EV) has been discovered that expresses higher in tumour tissue than adjacent normal tissue[47] and can capture cell-free DNA[48], which is essential in the tumour microenvironment, thus EV may be a biomarker in tumour senescence detection[49]. KEGG analysis also suggested that there is a difference in immune function between the two risk groups identified by the risk model, as immunity-related KEGG signalling pathways like "Proteoglycans in cancer", "Th17 cell differentiation" and "Primary immunodeficiency" were significantly enriched. To assess the expression level of the DEGs additionally, GSVA analysis[22] was conducted and the heatmap revealed the top 20 KEGG signalling pathways. The result of the analysis was consistent with the GO and KEGG analyses, suggesting change in immune function plays an important role in the prognosis of HNSCC. Several immune system–related KEGG signalling pathways, such as "Primary immunodeficiency", "T cell receptor signalling pathway", "Intestinal immune network for IgA production" and "B cell receptor signalling pathway", were found to be significantly suppressed in the high-risk group. Among them, "Primary immunodeficiency" was evaluated with the top absolute value of fold change, which was identified in an autography-related HNSCC prognosis model by Ren et al[50]. Interestingly, KEGG analysis strongly inhibits the upregulation of glycan biosynthesis and metabolism, indicating that active glycan metabolism levels may be a positive factor in the prognosis of HNSCC. This opinion was corroborated by previous studies[51-54].

Research in the present study indicated that immune system function influences the prognosis of HNSCC patients, suggesting the potential participation of the tumour microenvironment. The tumour microenvironment plays a pivotal role in tumour progression and metastasis. Its unique composition and subtype in individuals result in different therapeutic outcomes[55]. Thus, the evaluation of immune cell infiltration in the HNSCC tumour microenvironment was valuable. The IOBR package by R[29] was used to evaluate the tumour immune purity of each sample via ESTIMATE[27] and TIMER[28]. According to the ESTIMATE analysis, the HNSCC patients in the high-risk group presented with significantly lower immune scores and ESTIMATE scores, indicating that higher immune scores and ESTIMATE scores are distinct positive predictors of HNSCC prognosis, which is not consistent with other cancer types like lung adenocarcinoma[56], gastric cancer[57] and clear cell renal cell carcinoma[58]. Next, the infiltration level of six immune cells was evaluated via the TIMER algorithm. Significantly decreasing infiltration levels were noted in the high-risk group of each immune cell type except macrophage. In summary, the integrated evaluation suggested immunotherapy may benefit low-risk HNSCC patients, but may not be suitable for high-risk patients.

The present study has some limitations. First, the analysis was retrospective and dependent on publicly accessible databases, and it was challenging to account for the regional difference, which limited the application of the risk model in the clinic. Second, it is favourable to evaluate immune and stromal components systematically in the centre of the tumour and at the invasive margin given that the microenvironment may vary in each individual and tumour site. The various tumour regions could not be considered because the transcriptome profiles employed in the present investigation were all generated from a core sample of tumour tissue. Thus, there is a lack of consideration of the heterogeneity of different tumour sites within an individual, which may weaken the accuracy of the prognostic model to some extent. Therefore, further validation of cellular or animal models in the future will help to increase the credibility of the study. Meanwhile, a carefully planned, prospective, multicentre clinical trial that is conducted internationally is anticipated. Third, the authors used a GEO dataset for validation. Although having more data-

sets for validation would increase the persuasiveness of the findings, eligible validation sets are difficult to obtain. In the future, further tests on HNSCC clinical samples are needed to assess methylation site changes in HNSCC more comprehensively.

Conclusion

In conclusion, the present authors first established a 16-DNA methylation predictive HNSCC prognosis model associated with cellular senescence that may be utilised to select treatment for HNSCC patients. The selected 16 CpG sites that were identified as participating in cellular senescence were potential biofunction targets, revealing a novel research direction in the strategy for tumour therapy.

Acknowledgements

The analysis was supported by Sangerbox (Shanghai, China). The authors express their gratitude to the public databases TCGA and GEO.

Conflicts of interest

The authors declare no conflicts of interest related to this study.

Author contribution

Dr Ming Han YE contributed to the investigation, methodology, data interpretation, statistical analysis and manuscript draft; Dr Xin Yi HUANG contributed to the investigation and methodology; Drs Chun Jie LI, Qian Ju WU and Fei LIU contributed to the conceptualisation, project administration and supervision. All the authors approved the final version.

(Received May 11, 2023; accepted Aug 09, 2023)

References

1. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nat Rev Dis Primers 2020;6:92.
2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.CA Cancer J Clin 2021;71:209–249.
3. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576–582.
4. Chung CH, Parker JS, Karaca G, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 2004;5:489–500.
5. Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. Nat Rev Cancer 2018;18:269–282.
6. Bhat GR, Hyole RG, Li J. Head and neck cancer: Current challenges and future perspectives. Adv Cancer Res 2021;152:67–102.
7. Birch J, Gil J. Senescence and the SASP: Many therapeutic avenues. Genes Dev 2020;34:1565–1576.
8. Calcinotto A, Kohli J, Zagato E, Pellegrini L, Demaria M, Alimonti A. Cellular senescence: Aging, cancer, and injury. Physiol Rev 2019;99:1047–1078.
9. Coppé JP, Patil CK, Rodier F, et al. Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. PLoS Biol 2008;6:2853–2868.
10. Demirci D, Dayanc B, Mazi FA, Senturk S. The Jekyll and Hyde of cellular senescence in cancer. Cells 2021;10:208.
11. Schoetz U, Klein D, Hess J, et al. Early senescence and production of senescence-associated cytokines are major determinants of radioresistance in head-and-neck squamous cell carcinoma. Cell Death Dis 2021;12:1162.
12. Lee M, Nam HY, Kang HB, et al. Epigenetic regulation of p62/SQSTM1 overcomes the radioresistance of head and neck cancer cells via autophagy-dependent senescence induction. Cell Death Dis 2021;12:250.
13. Ahmadinejad F, Bos T, Hu B, et al. Senolytic-mediated elimination of head and neck tumor cells induced into senescence by Cisplatin. Mol Pharmacol 2022;101:168–180.
14. Klutstein M, Nejman D, Greenfield R, Cedar H. DNA methylation in cancer and aging. Cancer Res 2016;76:3446–3450.
15. Loughran O, Malliri A, Owens D, et al. Association of CDKN2A/p16INK4A with human head and neck keratinocyte replicative senescence: relationship of dysfunction to immortality and neoplasia. Oncogene 1996;13:561–568.
16. Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine learning-guided adjuvant treatment of head and neck cancer. JAMA Netw Open 2020;3:e2025881.
17. Wang J, Zhou CC, Sun HC, et al. Identification of several senescence-associated genes signature in head and neck squamous cell carcinoma. J Clin Lab Anal 2022;36:e24555.
18. Krishnan NM, Dhas K, Nair J, et al. A minimal DNA methylation signature in oral tongue squamous cell carcinoma links altered methylation with tumor attributes. Mol Cancer Res 2016;14:805–819.
19. Avelar RA, Ortega JG, Tacutu R, et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. Genome Biol 2020;21:91.
20. Liu Z, Zhao Q, Zuo ZX, et al. Systematic analysis of the aberrances and functional implications of ferroptosis in cancer. iScience 2020;23:101302.
21. Xu Y, Hong M, Kong D, Deng J, Zhong Z, Liang J. Ferroptosis-associated DNA methylation signature predicts overall survival in patients with head and neck squamous cell carcinoma. BMC Genomics 2022;23:63.
22. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
23. Gene Ontology Consortium, Blake JA, Dolan M, et al. Gene Ontology annotations and resources. Nucleic Acids Res 2013;41:D530–D535.

24. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017;45:D353–D361.
25. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
26. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739–1740.
27. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.
28. Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol 2016;17:174.
29. Zeng D, Ye Z, Shen R, et al. IOBR: Multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. Front Immunol 2021;12:687975.
30. Shibru B, Fey K, Fricke S, et al. Detection of immune checkpoint receptors - A current challenge in clinical flow cytometry. Front Immunol 2021;12:694055.
31. Schabath MB, Cote ML. Cancer progress and priorities: Lung cancer[J]. Cancer Epidemiol Biomarkers Prev 2019;28:1563–1579.
32. Almangush A, Mäkitie AA, Triantafyllou A, et al. Staging and grading of oral squamous cell carcinoma: An update. Oral Oncol 2020;107:104799.
33. Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. Exp Cell Res 1961;25:585–621.
34. Gorgoulis V, Adams PD, Alimonti A, et al. Cellular senescence: Defining a path forward. Cell 2019;179:813–827.
35. Di Micco R, Krizhanovsky V, Baker D, d'Adda di Fagagna F. Cellular senescence in ageing: From mechanisms to therapeutic opportunities. Nat Rev Mol Cell Biol 2021;22:75–95.
36. Prasanna PG, Citrin DE, Hildesheim J, et al. Therapy-induced senescence: Opportunities to improve anticancer therapy. J Natl Cancer Inst 2021;113:1285–1298.
37. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology 2013;38:23–38.
38. Ramu D, Shan TW, Hirpara JL, Pervaiz S. Cellular senescence: Silent operator and therapeutic target in cancer. IUBMB Life 2021;73:530–542.
39. Rao VK, Pal A, Taneja R. A drive in SUVs: From development to disease. Epigenetics 2017;12:177–186.
40. Brown R, Curry E, Magnani L, Wilhelm-Benartzi CS, Borley J. Poised epigenetic states and acquired drug resistance in cancer. Nat Rev Cancer 2014;14:747–753.
41. Majumder M, House R, Palanisamy N, et al. RNA-binding protein FXR1 regulates p21 and TERC RNA to bypass p53-mediated cellular senescence in OSCC. PLoS Genet 2016;12:e1006306.
42. Qie S, Majumder M, Mackiewicz K, et al. Fbxo4-mediated degradation of Fxr1 suppresses tumorigenesis in head and neck squamous cell carcinoma. Nat Commun 2017;8:1534.
43. Fan Y, Yue J, Xiao M, et al. FXR1 regulates transcription and is required for growth of human cancer cells with TP53/FXR2 homozygous deletion. Elife 2017;6:e26129.
44. Cao S, Zheng J, Liu X, et al. FXR1 promotes the malignant biological behavior of glioma cells via stabilizing MIR17HG. J Exp Clin Cancer Res 2019;38:37.
45. Adib E, Nassar AH, Akl EW, et al. CDKN2A alterations and response to immunotherapy in solid tumors. Clin Cancer Res 2021;27:4025–4035.
46. Baker DJ, Jin F, van Deursen JM. The yin and yang of the Cdkn2a locus in senescence and aging. Cell Cycle 2008;7:2795–2802.
47. Lázaro-Ibáñez E, Sanz-Garcia A, Visakorpi T, et al. Different gDNA content in the subpopulations of prostate cancer extracellular vesicles: Apoptotic bodies, microvesicles, and exosomes. Prostate 2014;74:1379–1390.
48. Kustanovich A, Schwartz R, Peretz T, Grinshpun A. Life and death of circulating cell-free DNA. Cancer Biol Ther 2019;20:1057–1067.
49. Ou HL, Hoffmann R, González-López C, Doherty GJ, Korkola JE, Muñoz-Espín D. Cellular senescence in cancer: From mechanisms to detection. Mol Oncol 2021;15:2634–2671.
50. Ren Z, Zhang L, Ding W, et al. Development and validation of a novel survival model for head and neck squamous cell carcinoma based on autophagy-related genes. Genomics 2021;113(1 Pt 2):1166–1175.
51. Zhao X, Brusadelli MG, Sauter S, et al. Lipidomic profiling links the Fanconi anemia pathway to glycosphingolipid metabolism in head and neck cancer cells. Clin Cancer Res 2018;24:2700–2709.
52. Reiss M, Maniglia CA, Sartorelli AC. Modulation of cell shedding and glycosaminoglycan synthesis of human malignant keratinocytes by all-trans-retinoic acid and hydrocortisone in vitro. J Invest Dermatol 1986;86:683–688.
53. Skandalis SS, Theocharis AD, Papageorgakopoulou N, Vynios DH, Theocharis DA. The increased accumulation of structurally modified versican and decorin is related with the progression of laryngeal cancer. Biochimie 2006;88:1135–1143.
54. Vynios DH, Theocharis DA, Papageorgakopoulou N, et al. Biochemical changes of extracellular proteoglycans in squamous cell laryngeal carcinoma. Connect Tissue Res 2008;49:239–243.
55. Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett 2017;387:61–68.
56. Wu J, Li L, Zhang H, et al. A risk model developed based on tumor microenvironment predicts overall survival and associates with tumor immunity of patients with lung adenocarcinoma. Oncogene 2021;40:4413–4424.
57. Wang H, Wu X, Chen Y. Stromal-immune score-based gene signature: A prognosis stratification tool in gastric cancer. Front Oncol 2019;9:1212.
58. Xu WH, Xu Y, Wang J, et al. Prognostic value and immune infiltration of novel signatures in clear cell renal cell carcinoma microenvironment. Aging (Albany NY) 2019;11:6999–7020.