Nonnegative Matrix Factorization Model-Based Construction For Molecular Clustering and Prognostic Assessment of Squamous Carcinoma of Head and Neck

Xin-Yu Li

Shanghai 9th Peoples Hospital Affiliated to Shanghai Jiaotong University School of Medicine

Xi-Tao Yang (✉ 118120@sh9hospital.org.cn)

Shanghai 9th Peoples Hospital Affiliated to Shanghai Jiaotong University School of Medicine

<https://orcid.org/0000-0003-4030-4394>

Primary research

Keywords: Nonnegative matrix factorization, prognostic, clustering, survival analysis

Posted Date: November 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1043798/v1>

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

[Read Full License](#)

Abstract

Purpose: Exploring nonnegative matrix factorization (NMF) model-based clustering and prognostic modeling of head and neck squamous carcinoma (HNSCC).

Methods: The transcriptome microarray data of HNSCC samples were downloaded from The Cancer Genome Atlas (TCGA) and Shanghai Ninth People's Hospital, and NMF clustering was constructed using the R software package. Relevant prognostic models were developed based on clustering.

Results: Based on NMF, all samples were divided into 2 subgroups. Predictive models were constructed by analysing the differential gene between the two subgroups. Results of survival analysis in the current study revealed that the high-risk group had a poor prognosis. Further, results of multi-factor Cox regression analysis revealed that the predictive model was an independent predictor of prognosis.

Conclusion: It was evident that the NMF-based prognostic model is a useful guide to the prognostic assessment of HNSCC.

Introduction

Squamous cell carcinoma of the head and neck (HNSCC) is one of the most common malignant tumours in the world. It accounts for more than 90% of all malignant tumours of the head and neck and thus poses a serious health risk to human beings^[1]. Treatment options for HNSCC are mainly based on TNM staging and a combination of surgical-based therapies (radiotherapy, chemotherapy and biotherapy)^[2]. Although, the majority of patients with HNSCC are presented with locally advanced disease with significant lymph node metastases their outcome has improved due to advances in multi-disciplinary treatment. However, the mortality rate is still above 55% and there is 40 to 60% recurrence and metastasis rate^[2-4]. Therefore, accurate prediction of the prognosis of patients with HNSCC is an important clinical guide.

Clustering is based on the principle that genes with similar expression patterns have similar or related functions. It is one of the important methods for processing gene expression data^[5]. Clustering is divided into one-way clustering and two-way clustering. One-way clustering is whereby only rows or columns are clustered and its results are more influenced by unrelated columns or rows. The commonly used one-way clustering algorithms include systematic clustering, self-organizing mapping clustering and principal component clustering. Two-way clustering is whereby the optimal set of sub-matrices are found in a matrix where the rows and columns are all significantly correlated. Furthermore, it allows overlap between classes, which is significant for gene chip data. Usually, a gene is not involved in only one biological process, but also each sample has multiple biological processes at the same time. Bidirectional clustering is thus more suitable for processing gene expression data. Nonnegative matrix factorization is a two-way clustering process^[6]. Nonnegative matrix factorization has 3 main advantages over other standard decomposition methods, namely, no parameters, good interpretability and good numerical

results^[6]. Nonnegative matrix factorization has been widely used for cancer classification based on gene expression profile data^[7].

The present study performed molecular clustering and prognostic modeling of HNSCC samples from TCGA database and validation group (collected by the Department of Oral and Maxillofacial Head and Neck Oncology, Shanghai Ninth People's Hospital) based on NMF. This was to appropriately classify the patients with HNSCC for treatment selection and prognosis predictions.

1.1 Data acquisition

The present study obtained RNA sequencing (RNA-seq) data of 502 HNSCC patients, 44 normal human head and neck samples as well as their corresponding clinical features from The Cancer Genome Atlas (TCGA) database ([https://portal.gdc.cancer.gov/ repository](https://portal.gdc.cancer.gov/repository)). Fresh HNSCC and normal tissues from 80 human patients were postoperatively collected from April 2010 to October 2016. In this procedure, frozen sections of tissue from the surgical margins were examined after completion of the extended resection. If the pathology is positive, continue to extend the excision until all margins are negative and remove normal tissue from the surrounding area.. A pathologist from the Department of Pathology of Shanghai Ninth People's Hospital diagnosed the patients in this study. Clinical data of the patients was as shown in the Tables 1 and 2. This study was approved by the Human Research Ethics Committee of the Ninth People's Hospital, Shanghai JiaoTong University School of Medicine (Shanghai, China). The informed consent was not available because this study was retrospective in nature.

1.2 Consensus Cluster of HNSCC samples based on the NMF model

Nonnegative matrix factorization cluster was constructed using the Consensus Cluster Plus package^[9]. Nonnegative matrix factorization hierarchical clustering is performed using the adjusted and unified dataset, the number of clusters k values are taken from 2 to 8. The value with better clustering stability is selected according to the clustering effect^[10]. In the present study Kaplan-Meier survival analysis was performed based on the results of NMF classification. To determine whether there is a significant difference in the survival prognosis of different groups of patients the different immune infiltration levels of each immune cell between the two groups were analyzed using the vioplot package in R software.

1.3 Construction of the prognostic model

For differential gene (DEGs) expression analysis was conducted using edgeR, where the threshold values were set to the absolute value of logFC greater than 1 and FDR less than 0.05 to screen for DEGs. DEG significantly associated with overall survival (OS) in patients with HNSCC were screened using univariate Cox regression analysis. Lasso regression analysis was then used to eliminate collinearity

between the genes. They were then included in a multifactorial Cox regression analysis model analysis for further screening. The final genes obtained were identified as predictive model component genes.

The prognostic signature was used as the risk score = $\sum_{i=1}^n \text{exp}_i * \beta_i$

Where, n, exp_i and β_i, represent the number of prognostic genes, the expression value and the coefficient of gene i, respectively. The risk score was calculated for each patient according to the formula, and the median of the scores was the cut-off value, whereas all the patients were divided into high-risk and low-risk groups using the cut-off value. Kaplan-Meier method was used to plot the overall survival curves for the different groups of patients and log-rank test was also performed.

The ROC curves and calibration plots were plotted to assess the predictive ability of the proposed model. The earlier described risk calculation formula was also used to calculate the risk score for each patient in validation group. The ROC and calibration plots were also used to validate the predictive ability of the model. The clinical specimens of 80 patients with HNSCC were collected during surgeries. The specimens were immersed into the RNA later solution (Invitrogen, USA) immediately and then used for RNA extraction or stored at -80 °C.

Total RNA was extracted from the fresh tissues using TRIzol reagent (Invitrogen) and cDNA was synthesized from 10 µg of total mRNA by using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturers' instructions. FastStart Universal SYBR Green Master Mix (Roche) and QuantStudio™ 6 Flex (Applied Biosystems) were used to perform qRT-PCR. Primers used for the qRT-PCR were as showed in Table S2. RNA-seq analysis was performed using the NovelBrain Cloud Analysis Platform, China. In brief, after total RNA was extracted, the cDNA libraries were constructed for each pooled RNA sample using the VAHTSTM Total RNAseq (H/M/R). The gene expression level was determined through the FPKM method.

1.4 Correlation analysis of model-independent prognostic and clinical characteristics

Univariate and multifactor Cox regression analyses of risk scores were performed to determine whether the model had independent prognostic value. If the risk score was significantly different from OS in both univariate and multivariate Cox analyses, the risk score was considered as an independent risk factor. Finally, DCA was introduced to prove the clinical validity of the model in the present study.

1.5 Immune infiltration analysis

The infiltrating score of 16 immune cells and the activity of 13 immune-related pathways were calculated using the single-sample gene set enrichment analysis (ssGSEA) method in the Gene Set Variation Analysis (GSVA) package of R software. The Benjamini–Hochberg (BH) method was used to

adjust the p values. Expression analysis was conducted to determine the relationship of risk score and immune-related genes, including m6a, ferroptosisiron death, cellular autophagy, tumor mutation burden (TMB) and major histocompatibility complex (MHC). Two hundred ninety eight patients treated with immunotherapy from the IMvigor210 were included to form an independent validation cohort and to verify the robustness of the classification and ability to predict the response to immunotherapy.

Results

2.1 Clustering based on the NMF model divides the samples into 2 subgroups

To reduce the effects of sample multicenter sources and batch processing, ComBat- data was corrected using the R package^[11]. Comprehensive judgment of clustering stability^[12, 13], stability was found to be better when k=2, thus k=2 was used for the judgement (Fig. 1a). Based on the survival curves and the log-rank test results (P<0.05), it was found that the prognosis of the 2 subgroups was significantly different (Fig. 1b-c). Further, it was found that the immune fine infiltration between the two subgroups was statistically significant different from each other (Fig. S1).

2.2 Construction of prognostic models

The above classification confirms a difference in prognosis between the two clusters and hence it warranted for further study. Therefore, the DEGs between the two clusters were subjected to univariate Cox analysis to obtain DEGs associated with prognosis, and the LASSO was also used to further screen the 13 associated genes (Fig. 1d). These genes were subjected to multifactorial Cox analysis and 9 DEGs were obtained as well as the correlation coefficients (Table 3). The prognostic model risk score was as follows: risk score= 0.78* expression level of *HAUS6*-0.38* expression level of *SCNN1D*+ 0.76* expression level of *S100A1*-1.64* expression level of *TNFRSF4*-1.51* expression level of *FBXO17*+ 0.75* expression level of *IRF9*+ 0.65* expression level of *IFI6*-0.41* expression level of *PTGS2*+ 0.32* expression level of *MSC*. The risk score of each patient was calculated based on the regression coefficients according to the prognostic model and the patients were divided into high-risk and low-risk groups by the median of the risk scores.

It was found that the survival of high-risk patients was significantly shorter compared with that of the patients in the low-risk group (Fig. 2a). The time-dependent ROC curves showed a 1-, 3- and 5- year time AUC of 0.852, 0.890 and 0.953, respectively which was in agreement with the calibration curves (Fig. 2b-c). Applying the same prognostic score to the validation set, the Kaplan-Meier survival curves showed that patients with high risk scores had lower OS than those with low risk scores, and the OS was significantly different between the two groups of patients (Fig. 3a). The 1- and -5-year AUC values in the validation set ranged from 0.767 to 0.862, indicating that the model still had good predictive performance in the external validation set (Fig. 3b).

Results of PCR in the present study showed that 9 genes still differed in the validation group (Fig. 3c). This was in line with the results of the TCGA cohort. Further, the results of Kaplan-Meier survival curve

analysis for the 9 genes of OS were as shown in Fig. S2. The above evaluation results suggest that the risk score model has good sensitivity and specificity for HNSCC prognosis prediction. A multifactorial Cox regression analysis was performed combining clinical indicators of the patients (risk score, age, gender, stage and grade among others). The results of this study showed that risk score were associated with survival (Table 4). Second, decision curves were made to determine the clinical net benefit derived from the use of the model. Decision curve analysis demonstrated that the model was clinically useful (Fig. 3d). In conclusion, the risk score can be used as a prognostic indicator independent of other clinical factors for the prognosis of HNSCC.

2.3 Immunogenesis and enrichment analysis

Analysis of the relationship between risk score and m6a, ferroptosis, cellular autophagy and other related genes revealed that the risk score was closely related to immune-related genes (Fig. S3). The TMB refers to the number of base mutations per million bases and is a marker for the efficacy of immune checkpoint inhibitors. The higher the TMB, the more neoantigens can be recognized by T cells at the end and the better the immunotherapy effect. Our analysis showed a negative correlation between risk score and TMB, which could also be the reason for the poor prognosis of high-risk patients (Fig. 4a). It was found that there was a higher probability of higher benefit for high-risk patients in immunotherapy treatment (complete response (CR), partial response (PR), no clinical benefit (progressive disease (PD) or Stable Disease (SD)). This provides new options for the treatment of patients with subsequent tumors (Fig. 4b). According to the results of GSEA, the high-risk group was enriched in dilated cardiomyopathy whereas the low-risk group was mainly associated with tumorigenesis development. The GSEA analysis partially explained the biological differences between the low- and high-risk groups at the genetic and pathway level (Fig. 4c). Enrichment and signaling pathway analysis were performed for DEGs and it was hence possible to understand the biological functions of DEGs. The DEGs were mainly enriched in cell cycle, DNA replication, catalytic activity, acting on DNA, chromosomal region, human papillomavirus infection, organelle fission and condensed chromosome (Fig. 4d). We further compared the enrichment scores of 16 types of immune cells and the activity of 13 immune-related pathway between the low and high-risk groups by employing the ssGSEA. The high-risk subgroup generally had higher levels of infiltration of immune cells especially of T helper cells, Macrophages, regulatory T (Treg) cells and tumour-infiltrating lymphocytes (TILs) (Fig. 5a). Except for the MHC_class_I and type-1 IFN response pathway, the other 11 immune pathways showed lower activity in the high-risk group than in the low-risk group in the TCGA cohort (Fig. 5b).

2.4 Drug sensitivity analysis

The highest negative correlation score is for chrysin (-0.776). Chrysin is a drug which has a variety of biological activities, such as anti-tumor, anti-inflammatory, anti-bacterial, anti-anxiety and anti-oxidant pharmacological activities^[14], hence suggesting a possible therapeutic effect in HNSCC. The next highest scores was MS-275. Previous studies have shown that it has a selective killing effect on gastric

adenocarcinoma cells)^[15], 1, 4-chrysenequinone (an Ahr-activator) and piperlongumine (inhibits tumor autophagy leading to reduced cell proliferation viability) (Table S1).

Discussion

Head and neck squamous carcinoma is one of the more common malignancies worldwide, with 550,000 new cases and about 380,000 deaths per year^[16, 17]. It is not only aggressive and lethal, but also causes serious facial deformities, speech, chewing and swallowing dysfunctions as well as psychosocial problems to patients. Although surgical radical techniques, repair and reconstruction techniques for HNSCC have become increasingly sophisticated, their 5-year survival rate has not improved significantly in the last 20 years. The first research results on NMF algorithms was publication in 1999. In 2003 Kim first used NMF for clustering of genes and identifying subsystems of functional cells^[18].

To better predict the prognosis of HNSCC, this study staged patient with HNSCC s into two subgroups based on the NMF model. It was found that there was significant differences in OS between the two subgroups, with patients in the subgroup with more abundant immune cell infiltration having better prognostic indicators. This suggests that associated immune cells can influence the prognosis of patients with HNSCC. The enrichment analysis of differential genes between the two groups also revealed the existence of enrichment in immune-related functions.

A prognostic risk model was constructed consisting of nine genes. The scores of patients were calculated based on the risk model and divided into two groups of high and low risk. Further, it was found that there was had a significant difference in the prognosis of patients in the two groups of high and low risk whereas the prognosis of high-risk patients was significantly lower than that of the low-risk patients. The ROC curve and calibration curve of the model also achieved remarkable results which revealed that the model has better discriminatory ability. The DCA also demonstrated that the reliability and accuracy of the prediction model was better than the other clinical indicators.

Tumour necrosis factor receptor superfamily, member4 (TNFRSF4) is one of the component genes of the model and is predominantly inducibly expressed in activated CD4+ and CD8+ cells^[19]. Previous studies have shown that binding of TNFRSF4 to ligands promotes clonal proliferation of T cells, enhances T cell memory, proliferation, immune surveillance and killer cells, as well as prevents the development of immune tolerance^[20]. In addition, TNFRSF4 expressing positive T cells can reduce suppressive factors in the tumor immune microenvironment and effectively inhibit tumor invasion and metastasis^[21]. The distribution of TNFRSF4 expression in breast cancer, melanoma, and lymphoma has been discussed in the previous studies^[22, 23], and targeted TNFRSF4 treatment can significantly play a role in anti-breast cancer and melanoma^[24]. In the present study, TNFRSF4 was a protective factor for HNSCC which was in agreement with previous studies.

Enrichment analysis of the differential genes revealed that the genes were mainly enriched in cell cycle, DNA replication, catalytic activity and acting on DNA. The GSEA analysis also revealed that the low-risk

patients were predominantly enriched in immunodeficiency and tumor-associated pathways. Further, the analysis of immune cell infiltration in both groups of patients revealed a positive correlation between the macrophages and risk scores. This phenomenon was associated with the fact that macrophages are essential in the HNSCCC microenvironment for the regulation of the inflammatory reactions^[17]. It was hence suggested that the model may serve as a predictor of increased immune cell infiltration.

Several studies have reported that increase of macrophages is associated with poor cancer prognosis^[25]. Further, the macrophage infiltration in the tumor microenvironment can promote tumor growth, angiogenesis, invasion and metastasis^[26]. The present study also analyzed the correlation between risk scores and immune-related genes. It was revealed that the risk scores were strongly associated with m6a-related, iron death-related and autophagy-related genes. Recently, the development of tumor immunotherapy has rapidly progressed and become of a great research interest^[27]. Immune-suppressants such as PD-1/PD-L1 suppressants have successfully emerged in the recent past^[28]. The present study analyzed the responsiveness to PD-1/PD-L1 in both high and low risk groups of patients and it was evident that high risk patients responded to immunotherapy better than the low risk patients.

Conclusions

In conclusion, based NMF algorithm, this study screened for DEGs and constructed an associated prognostic model which could independently predict prognosis in patients with HNSCC. The predictive performance of the model was found to be stable and can assist in providing reference for individualized treatment of patients with HNSCC. Further, the genes in prognostic risk models also provided new therapeutic targets for the exploration of immunotherapy for HNSCC.

Declarations

Competing interests: no conflicts of interest.

Contributors: Xi-tao Yang designed experiments; Xin-yu Li carried out experiments, and wrote the manuscript, Xi-tao Yang performed manuscript review.

Category: original article

Availability of data and material Additional data and material are available upon request

Consent for publication: All authors consent for publication.

Financial Support: This study received Fundamental research program funding of Ninth People's Hospital affiliated to Shanghai Jiao Tong university School of Medicine (No. JYZZ076), Clinical Research Program of Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine (No. JYLJ201801, JYLJ201911), the China Postdoctoral Science Foundation (No. 2017M611585) and the National Natural Science Foundation of China (No. 81871458).

Patient consent: Obtained.

Ethics approval: All procedures performed in the studies involving human participants were in accordance with the Scientific Research Projects Approval Determination of Independent Ethics Committee of The First Affiliated Hospital of Zhengzhou University, with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Acknowledgements: Home for Researchers editorial team www.home-for-researchers.com

References

1. Stashenko P, Yost S, Choi Y, et al. The Oral Mouse Microbiome Promotes Tumorigenesis in Oral Squamous Cell Carcinoma[J]. *mSystems*,2019,4(4):e319-e323.DOI:10.1128/mSystems.00323-19
2. Solomon B, Young R J, Rischin D. Head and neck squamous cell carcinoma: Genomics and emerging biomarkers for immunomodulatory cancer treatments.[J]. *Seminars in cancer biology*,2018,52(Pt 2):228–240.DOI:10.1016/j.semcancer.2018.01.008
3. Pan M, Schinke H, Luxenburger E, et al. EpCAM ectodomain EpEX is a ligand of EGFR that counteracts EGF-mediated epithelial-mesenchymal transition through modulation of phospho-ERK1/2 in head and neck cancers[J]. *PLoS biology*,2018,16(9):e2006624.DOI:10.1371/journal.pbio.2006624
4. Yang H, Cao Y, Li Z, et al. The role of protein p16(INK4a) in non-oro-pharyngeal head and neck squamous cell carcinoma in Southern China.[J]. *Oncology letters*,2018,16(5):6147–6155.DOI:10.3892/ol.2018.9353
5. Leite Pereira A, Tchitchek N, Lambotte O, et al. Characterization of Leukocytes From HIV-ART Patients Using Combined Cytometric Profiles of 72 Cell Markers.[J]. *Frontiers in immunology*,2019,10:1777.DOI:10.3389/fimmu.2019.01777
6. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization.[J]. *BMC bioinformatics*,2010,11:367.DOI:10.1186/1471-2105-11-367
7. Brunet J, Tamayo P, Golub T R, et al. Metagenes and molecular pattern discovery using matrix factorization.[J]. *Proceedings of the National Academy of Sciences of the United States of America*,2004,101(12):4164–4169
8. Shen Y, Li X, Wang D, et al. Novel prognostic model established for patients with head and neck squamous cell carcinoma based on pyroptosis-related genes.[J]. *Translational oncology*,2021,14(12):101233.DOI:10.1016/j.tranon.2021.101233
9. Jiang M, Kang Y, Sewastianik T, et al. BCL9 provides multi-cellular communication properties in colorectal cancer by interacting with paraspeckle proteins[J]. *Nature communications*,2020,11(1):19.DOI:10.1038/s41467-019-13842-7
10. Tandon A, Albeshri A, Thayananthan V, et al. Fast consensus clustering in complex networks.[J]. *Physical review. E*,2019,99(4-1):42301.DOI:10.1103/PhysRevE.99.042301

11. Johnson W E, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods.[J]. *Biostatistics* (Oxford, England),2007,8(1):118–127
12. Sadanandam A, Lyssiotis C A, Homiczko K, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy.[J]. *Nature medicine*,2013,19(5):619–625.DOI:10.1038/nm.3175
13. Verhaak R G W, Hoadley K A, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. [J]. *Cancer cell*,2010,17(1):98–110.DOI:10.1016/j.ccr.2009.12.020
14. Mani R, Natesan V. Chrysin: Sources, beneficial pharmacological activities, and molecular mechanism of action.[J]. *Phytochemistry*,2018,145:187-196.DOI:10.1016/j.phytochem.2017.09.016
15. Zhang Y, Adachi M, Zhao X, et al. Histone deacetylase inhibitors FK228, N-(2-aminophenyl)-4-[N-(pyridin-3-yl-methoxycarbonyl)amino- methyl]benzamide and m-carboxycinnamic acid bis-hydroxamide augment radiation-induced cell death in gastrointestinal adenocarcinoma cells.[J]. *International journal of cancer*,2004,110(2):301–308
16. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016.[J]. *Lancet* (London, England),2017,390(10100):1211-1259.DOI:10.1016/S0140-6736(17)32154-2
17. Shen Y, Li X, Wang D, et al. Novel prognostic model established for patients with head and neck squamous cell carcinoma based on pyroptosis-related genes.[J]. *Translational oncology*,2021,14(12):101233.DOI:10.1016/j.tranon.2021.101233
18. Kim P M, Tidor B. Subsystem identification through dimensionality reduction of large-scale gene expression data.[J]. *Genome research*,2003,13(7):1706–1718
19. Aspeslagh S, Postel-Vinay S, Rusakiewicz S, et al. Rationale for anti-OX40 cancer immunotherapy.[J]. *European journal of cancer* (Oxford, England: 1990),2016,52:50–66.DOI:10.1016/j.ejca.2015.08.021
20. Buchan S L, Rogel A, Al-Shamkhani A. The immunobiology of CD27 and OX40 and their potential as targets for cancer immunotherapy.[J]. *Blood*,2018,131(1):39–48.DOI:10.1182/blood-2017-07-741025
21. Bell R B, Leidner R S, Crittenden M R, et al. OX40 signaling in head and neck squamous cell carcinoma: Overcoming immunosuppression in the tumor microenvironment.[J]. *Oral oncology*,2016,52:1-10.DOI:10.1016/j.oraloncology.2015.11.009
22. Marabelle A, Kohrt H, Sagiv-Barfi I, et al. Depleting tumor-specific Tregs at a single site eradicates disseminated tumors.[J]. *The Journal of clinical investigation*,2013,123(6):2447–2463
23. Xie F, Wang Q, Chen Y, et al. Costimulatory molecule OX40/OX40L expression in ductal carcinoma in situ and invasive ductal carcinoma of breast: an immunohistochemistry-based pilot study.[J]. *Pathology, research and practice*,2010,206(11):735–739.DOI:10.1016/j.prp.2010.05.016
24. Weinberg A D, Rivera M M, Prell R, et al. Engagement of the OX-40 receptor in vivo enhances antitumor immunity.[J]. *Journal of immunology* (Baltimore, Md.: 1950),2000,164(4):2160–2169

25. Tian Z, Hou X, Liu W, et al. Macrophages and hepatocellular carcinoma.[J]. Cell & bioscience,2019,9:79.DOI:10.1186/s13578-019-0342-7
26. Kim J, Bae J. Tumor-Associated Macrophages and Neutrophils in Tumor Microenvironment.[J]. Mediators of inflammation,2016,2016:6058147.DOI:10.1155/2016/6058147
27. Wang Y, Liu Y, Du X, et al. The Anti-Cancer Mechanisms of Berberine: A Review.[J]. Cancer management and research,2020,12:695–702.DOI:10.2147/CMAR.S242329
28. Han R, Xiao Y, Yang Q, et al. Ag(2)S nanoparticle-mediated multiple ablations reinvigorates the immune response for enhanced cancer photo-immunotherapy.[J]. Biomaterials,2021,264:120451.DOI:10.1016/j.biomaterials.2020.120451

Tables

Table. 1: The basic clinical characteristics of derivation cohort

Characteristic	levels	Overall
n		502
T stage, n (%)	T1	33 (6.8%)
	T2	144 (29.6%)
	T3	131 (26.9%)
	T4	179 (36.8%)
N stage, n (%)	N0	239 (49.8%)
	N1	80 (16.7%)
	N2	154 (32.1%)
	N3	7 (1.5%)
M stage, n (%)	M0	472 (99%)
	M1	5 (1%)
Clinical stage, n (%)	Stage I	19 (3.9%)
	Stage II	95 (19.5%)
	Stage III	102 (20.9%)
	Stage IV	272 (55.7%)
Gender, n (%)	Female	134 (26.7%)
	Male	368 (73.3%)
Age, n (%)	<=60	245 (48.9%)
	>60	256 (51.1%)
Race, n (%)	Asian	10 (2.1%)
	Black or African American	47 (9.7%)
	White	428 (88.2%)
Age, median (IQR)		61 (53, 69)

Table. 2: The basic clinical characteristics of validation cohort

Characteristic	levels	Overall
n		80
Pathologic T stage, n (%)	T2	14 (17.5%)
	T3	32 (40%)
	T4	34 (42.5%)
Pathologic N stage, n (%)	N0	40 (50%)
	N1	16 (20%)
	N2	20 (25%)
	N3	4(5%)
Pathologic M stage, n (%)	M0	51 (65.4%)
	M1	27 (34.6%)
Pathologic stage, n (%)	Stage II	39 (49.4%)
	Stage III	36 (45.6%)
	Stage IV	4 (5.1%)
Gender, n (%)	Female	35 (43.8%)
	Male	45 (56.2%)
Age, n (%)	<=60	40 (50%)
	>60	40 (50%)
Age, median (IQR)		61.5 (51, 74.25)

Table. 3: Gene correspondence coefficient

id	coef
HAUS6	0.781261442
SCNN1D	-0.382482986
S100A1	0.760949529
TNFRSF4	-1.642657948
FBXO17	-1.512965493
IRF9	0.751127416
IFI6	0.654360943
PTGS2	-0.41098144
MSC	0.322998811

Table. 4: Univariate and multivariate Cox regression model were performed to detect the prognostic elements.

Characteristics	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
T stage				
T1	Reference			
T2	1.086 (0.568-2.074)	0.803		
T3	1.461 (0.769-2.773)	0.247		
T4	1.249 (0.665-2.344)	0.490		
N stage				
N0	Reference			
N1	1.058 (0.728-1.539)	0.768	0.999 (0.682-1.465)	0.997
N2&N3	1.404 (1.038-1.900)	0.028	1.469 (1.077-2.003)	0.015
M stage				
M0	Reference			
M1	4.745 (1.748-12.883)	0.002	4.288 (1.563-11.761)	0.005
Age				
<=60	Reference			
>60	1.252 (0.956-1.639)	0.102		
Gender				
Female	Reference			
Male	0.764 (0.574-1.018)	0.066	0.779 (0.579-1.046)	0.097
Riskscore(low vs high)	0.770 (0.672-0.883)	<0.001	0.757 (0.660-0.870)	<0.001
Clinical stage				
Stage I&Stage II	Reference			
Stage III&Stage IV	1.217 (0.878-1.688)	0.238		

Figures

Figure 1

a Nonnegative matrix factorization cluster analysis. The best fitted cluster was $k = 2$ value. KM curves showing PFS (b) and OS (c) for 2 cluster. d. Tenfold cross-validated error (first vertical line equals the minimum error, whereas the second vertical line shows the cross-validated error within 1 standard error of the minimum) (left). The profile of coefficients in the model at varying levels of penalization plotted against the log (lambda) sequence(right).

Figure 2

Prognostic analysis of the model in the derivation cohort. a. AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the derivation cohort. b. Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the derivation cohort. c. Calibration plot for model.

Figure 3

Validation of the model in the validation cohort. a. Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the validation cohort. b. AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the validation cohort. c. Results of qRT-PCR analysis. d. The decision curve analyses (DCA) for the clinical values of this model.

Figure 5

Comparison of the ssGSEA scores between different risk groups in the derivation cohort. The scores of 16 immune cells (a) and 13 immune-related functions (b) are displayed in boxplots. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns = not significant

Supplementary Files

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

- [TableS1.xls](#)
- [TableS2.xlsx](#)
- [s1.tif](#)
- [s2.tif](#)
- [s3.tif](#)