Association between cancer stem cell gene signature and prognosis in head and neck squamous cell carcinoma

Su Il Kim
Kyung Hee University Medical Center
Seon Rang Woo
Kyung Hee University Medical Center
Joo Kyung Noh
Kyung Hee University
Min Kyeong Lee
Kyung Hee University
Young Chan Lee
Kyung Hee University Medical Center
Jung Woo Lee
Kyung Hee University
Moonkyoo Kong
Kyung Hee University Medical Center
Seong-Gyu Ko
Kyung Hee University
Young-Gyu Eun (ygeun@hanmail.net)
Kyung Hee University Medical Center

Research Article

Keywords: head and neck squamous cell carcinoma, cancer stem cell, gene signature, overall survival, recurrence-free survival

Posted Date: April 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1497900/v1

Abstract

Background: Various cancer stem cell (CSC) biomarkers and genes encoding them in head and neck squamous cell carcinoma (HNSCC) have been identified and evaluated. However, the validity of these factors in the prognosis of HNSCC has been questioned and remains to be ascertained. In this study, we examined the clinical significance of CSC biomarker genes in HNSCC by using five publicly available HNSCC cohorts.

Methods: To predict the prognosis of patients with HNSCC, we developed and validated CSC gene signatures whose mRNA expression was correlated with at least one of four CSC genes, CD44, MET, ALDH1A1, and BMI1.

Results: Patients in TCGA HNSCC cohort were classified into CSC gene-associated high-risk (CSC-HR; n=285) and CSC gene-associated low-risk (CSC-LR; n=281) subgroups. The 5-year overall survival and recurrence-free survival rates were significantly lower in the CSC-HR subgroup than in the CSC-LR subgroup (p=0.04 and 0.02, respectively). The clinical significance of the CSC gene signature was validated using four independent cohorts. Cox proportional hazards models showed that the CSC gene signature was an independent prognostic factor for patients with HNSCC. Furthermore, CSC gene signature was associated with the prognosis of patients with HNSCC who received radiotherapy or those with HPV (-) status.

Conclusion: The CSC gene signature was associated with the prognosis of HNSCC and may help in personalized treatments for patients with HNSCC.

Background

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide, and includes all cancers that occur in the mucosa of the oropharynx, oral cavity, hypopharynx, or larynx [1]. Approximately 650,000 new cases of HNSCC occur and 350,000 patients die worldwide every year [2]. Despite advances in therapeutic methods, the survival rates of this condition have not markedly improved over the past few decades [3].

HPV status is a well-known factor that influences the prognosis of patients with HNSCC [4]. Recently, various molecular markers that influence HNSCC prognosis have been actively identified for precision medicine and personalized treatment [5]. However, there are still many other molecular markers that require further investigation. Cancer stem cells (CSCs) and CSC markers are considered important targets in this context [6].

CSCs make up part of the tumor that has long-term repopulation potential, the ability to evade cell death, clonal tumor initiation capacity, and self-renewal properties [7]. CSCs have been identified through the

expression of cell surface markers, which are mostly selected by embryonic stem cells or by self-properties involved with tissue development lineage molecules [8]. The most well-known CSC marker is CD44, which is associated with cell proliferation, angiogenesis, adhesion, and migration in tumorigenesis [9]. In addition to CD44, 27 CSC biomarkers have been reported and evaluated in HNSCC [10].

However, CSCs are composed of only a small proportion of cancer cells [11] and can be regulated by external forces and cell-autonomous forces [12]. In other words, CSCs may not necessarily be rare within tumors, and non-CSCs in tumors can be reversibly reprogrammed to become CSCs [13]. Thus, genes encoding CSC markers in tumors need to be analyzed, regardless of whether they are CSCs or non-CSCs. However, the validity and clinical significance of these genes has been questioned recently and remains to be ascertained in HNSCC. Additionally, the validity and underlying relationship among these CSC biomarker genes to predict the prognosis of HNSCC have not been demonstrated. Therefore, further research is needed to validate the role of CSC biomarker genes in HNSCC and to determine personalized treatments for patients with HNSCC.

In this study, we analyzed the genomic data of patients with HNSCC to determine the molecular subtypes associated with CSC biomarker genes, thereby predicting their prognosis. We hypothesized that the investigation of mRNA expression of various genes including CSC biomarker genes in The Cancer Genome Atlas (TCGA) HNSCC cohort would generate CSC gene-related molecular signatures, which could be validated in various independent HNSCC cohorts. We also investigated the prognostic importance of CSC gene signatures in various subgroups of patients with HNSCC.

Methods

Patient cohorts

Selection of patient cohorts was referred form our previous study [14]. Gene expression levels and clinical data from five independent cohorts were downloaded from public databases. Using the UCSC Cancer Genomics Browser (http://xenabrowser.net/datapages/), clinical and gene expression data of TCGA cohort (n = 566) were obtained. Using the National Center for Biotechnology Information Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo), corresponding data from the Institute for Medical Informatics, Statistics and Epidemiology (Leipzig cohort, GSE65858, n = 270) [15], Fred Hutchinson Cancer Research Center (FHCRC cohort, GSE41613, n = 97) [16], MD Anderson Cancer Center (MDACC cohort, GSE42743, n = 74) [16], and AHEPA Hospital in Thessaloniki (Greece cohort, GSE27020, n = 109) [17] were obtained. The gene expression profile of TCGA cohort was measured using Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA), while that of the Leipzig cohort was measured using Illumina HumanHT-12 v4.0 Expression Beadchip, and those of the FHCRC, MDACC, and Greece cohorts were measured using Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix Inc., California, USA). All gene expression data were standardized using different platforms.

Selection of reference CSC genes (training cohort)

Search results of reported CSC biomarkers and encoding genes were obtained from Xiao. et al [10]. Specifically, CSC biomarkers and encoding genes were searched in PubMed using the terms ‘tumor stem cells’, ‘tumor stem-like cells’, ‘CSCs’, ‘cancer stem cells’, ‘cancer stem-like cells’, and ‘HNSCC’, ‘head and neck squamous cell carcinoma’. Twenty-eight genes were selected from studies that met the following criteria: 1) studies in humans, 2) studies showing validated evidence, and 3) an original research paper. Case reports, comments, reviews, letters to the editor, and conference abstracts were excluded.

To select reference CSC genes among the 28 CSC genes, we additionally searched articles about CSC biomarkers or encoding genes in HNSCC written between 2012 and 2021. Articles were searched in PubMed using the terms ‘HNSCC’, ‘head and neck squamous cell carcinoma’, and each term about ‘CSC biomarkers and encoding genes’. We selected genes satisfying the following criteria: (a) CSC biomarkers showing clinical significance when classified according to expression and (b) those that have been studied more than twice.

The selected genes were analyzed using a training cohort (TCGA cohort). First, TCGA cohorts were classified into two subgroups according to the mRNA expression of each gene. To define dichotomous cut-off values for continuous mRNA expression for each gene, an online tool (http://molpath.charite.de/cutoff/) was applied [18]. The Kaplan–Meier method was used to generate survival curves for each subgroup for each gene. The log-rank test was used to compare the prognoses of the two subgroups for each gene. CSC genes showing significant differences in the 5-year overall survival (OS) or recurrence-free survival (RFS) between the two subgroups classified according to the mRNA expression were selected as reference CSC genes.

Development of CSC gene-related signature

Gene expression data were analyzed using TCGA cohort to identify CSC gene-related signatures in HNSCC. Genes whose mRNA expression levels were negatively or positively correlated with at least one of the CSC gene markers were selected. We then performed hierarchical clustering analysis with the centered correlation coefficient as a measure of similarity and a complete linkage clustering method using the Gene Cluster 3.0 program [19]. Patients were divided into CSC gene-associated high-risk (CSC-HR) and low-risk (CSC-LR) subgroups. The subgroup showing significantly lower survival rates than the other group was defined as the CSC-HR subgroup. The Java Treeview program was used to generate heat maps for cluster analysis.

Construction of prediction models and validation in the four independent cohorts

Before constructing the prediction models, all gene expression data for each cohort were standardized because they were generated using different platforms. The Support Vector Machine (SVM) class prediction engine was used to test the ability of CSC gene-related signatures to predict the class of patients with HNSCC in another four independent cohorts [20]. Gene expression data in TCGA cohort were combined to form a series of classifiers according to the SVM algorithm, and the robustness of the classifier was estimated according to the misclassification rate determined during leave-one-out cross-

validation of the training set using BRB–Array Tools [21]. Validation was conducted in four independent cohorts (Leipzig, FHCRC, MDACC, and Greece).

Pathway analysis

To identify gene ontology categories with significantly enriched gene numbers, 81 CSC gene-related signatures were analyzed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (version 6.8) [22]. To map the CSC gene signature to the reference set of direct and indirect relationships, default settings from the software were utilized. Then, relevant inputs to the gene list, such as the biological functions and molecular networks, were generated using the software's algorithm. Significant gene annotation was determined using a two-tailed Fisher's exact test ($p < 0.05$).

Statistical analysis

Gene expression data and survival data were used to test prognostic significance. Statistical analysis of data was also referred from our previous study [14]. OS was defined as the number of months from the date of diagnosis to death. The number of months from the date of diagnosis to recurrence was defined as RFS. The Kaplan–Meier method was used to produce OS and RFS curves in each subgroup of each cohort. The log-rank test was used to compare OS and RFS between each subgroup. Univariate and multivariate Cox regression models were used to evaluate independent prognostic factors associated with the survival of patients with HNSCC. The results of the Cox regression analyses were reported as hazard ratios (HRs), 95% confidence intervals (95% CIs), and p values. The R software package (http://www.r-project.org) was used for all statistical analyses. Statistical significance was set at $p < 0.05$.

Results

Development of CSC gene-related signature in patients with HNSCC

Among 28 CSC genes encoding CSC biomarker proteins validated in HNSCC [10], we selected seven CSC genes—CD44, MET, ALDH1A1, BMI1, PROM1, SOX2, and POU5F1– from a literature search that satisfied the following criteria: (a) showing clinical significance associated with expression of corresponding CSC biomarker proteins in HNSCC and (b) studied more than twice over the past 10 years [5, 8, 9, 23–31]. In TCGA cohort, high expression of four CSC genes (CD44, MET, ALDH1A1, and BMI1) was significantly associated with patient prognosis ($p = 0.0069$, 0.0051, 0.028, and 0.021, respectively; Fig. S1). Thus, these four genes were selected as reference CSC genes.

We identified the genes whose mRNA expression was correlated with at least one of the four reference CSC genes in TCGA cohort. A total of 81 genes were identified and were selected as CSC gene-related signatures (Fig. S2A, Table S1). Using the CSC gene signature, patients in TCGA cohort (n = 566) were divided into the CSC-HR (n = 285) and CSC-LR (n = 281) subgroups (Fig. 1A). The CSC-HR subgroup showed significantly higher mRNA expression of CD44 and MET than the CSC-LR subgroup ($p < 0.0001$ and $p < 0.0001$, respectively). And, the CSC-HR subgroup showed significantly lower mRNA expression of

ALDH1A1 and BMI1 than the CSC-LR subgroup ($p < 0.0001$ and $p < 0.0001$, respectively). Using Kaplan–Meier analysis and log-rank test, we confirmed that the CSC-HR subgroup showed significantly lower 5-year OS and RFS rates than the CSC-LR subgroup ($p = 0.04$ and $p = 0.02$, respectively; Fig. 1B-C).

Independent validation of CSC gene-related signature

The CSC gene signature was validated using four independent cohorts: the Leipzig (n = 270), FHCRC (n = 97), MDACC (n = 74), and Greece (n = 109) cohorts (Fig. S2B). The details of the clinical and pathological characteristics of each cohort used in this study are shown in Table 1. Patients in each validation cohort were efficiently classified into the CSC-HR and CSC-LR subgroups based on the CSC gene signature. The CSC-LR subgroup in each validation cohort showed better prognosis than the CSC-HR subgroup (Fig. 2). In the Leipzig cohort, the CSC-HR subgroup showed lower 5-year OS rates than the CSC-LR subgroup ($p = 0.06$; Fig. 2A). Furthermore, the 5-year OS rates were significantly lower in the CSC-HR subgroups than in the CSC-LR subgroups in the FHCRC and MDACC cohorts ($p < 0.0001$ and $p = 0.02$, respectively; Fig. 2B-C). Also, in the Greece cohort, the CSC-HR subgroup showed significantly lower 5-year RFS rates than the CSC-LR subgroup ($p = 0.009$; Fig. 2D).

Table 1
Clinical and pathological characteristics of the five independent HNSCC cohorts.

Characteristics	TCGA cohort (n = 566)	Leipzig cohort (n = 270)	FHCRC cohort (n = 97)	MDACC cohort (n = 74)	Greece cohort (n = 109)
Age					
≥ 60	316 (55.83%)	117 (43.33%)	47 (48.45%)	37 (50.00%)	74 (67.89%)
< 60	249 (43.99%)	153 (56.67%)	50 (51.55%)	37 (50.00%)	35 (32.11%)
Unknown	1 (0.18%)	0	0	0	0
Sex					
Male	415 (73.32%)	223 (82.59%)	66 (68.04%)	58 (78.38%)	104 (95.41%)
Female	151 (26.68%)	47 (17.41%)	31 (31.96%)	16 (21.62%)	5 (4.59%)
Smoking					
Yes	423 (74.73%)	222 (82.22%)	NA	59 (79.73%)	108 (99.08%)
No	128 (22.61%)	48 (17.78%)	NA	15 (20.27%)	1 (0.92%)
Unknown	15 (2.65%)	0	NA	0	0
Alcohol					
Yes	371 (65.55%)	239 (88.52%)	NA	NA	58 (53.21%)
No	182 (32.16%)	31 (11.48%)	NA	NA	51 (46.79%)
Unknown	13 (2.3%)	0	NA	NA	0
Tumor site					
Oral cavity	346 (61.13%)	83 (30.74%)	97 (100%)	71 (95.95%)	0
Oropharynx	82 (14.49%)	102 (37.78%)	0	3 (4.05%)	0

HNSCC, head and neck squamous cell carcinoma; TCGA, The Cancer Genome Atlas; FHCRC, Fred Hutchinson Cancer Research Center; MDACC, MD Anderson Cancer Center; CSC, cancer stem cell; CSC-HR, CSC gene-associated high-risk; CSC-LR, CSC gene-associated low-risk; NA, not available

Characteristics	TCGA cohort (n = 566)	Leipzig cohort (n = 270)	FHCRC cohort (n = 97)	MDACC cohort (n = 74)	Greece cohort (n = 109)
Larynx	128 (22.61%)	48 (17.78%)	0	0	109 (100%)
Hypopharynx	10 (1.77%)	33 (12.22%)	0	0	0
Unknown	0	4 (1.48%)	0	0	0
T classification					
T1-T2	218 (38.52%)	115 (42.59%)	NA	30 (40.54%)	NA
T3-T4	344 (60.78%)	155 (57.41%)	NA	44 (59.46%)	NA
Unknown	4 (0.71%)	0	NA	0	NA
N classification					
Negative	295 (52.12%)	94 (34.81%)	NA	42 (56.76%)	NA
Positive	267 (47.17%)	176 (65.19%)	NA	32 (43.24%)	NA
Unknown	4 (0.71%)	0	NA	0	NA
Stage					
I-II	135 (23.85%)	55 (20.37%)	41 (42.27%)	19 (25.68%)	30 (27.52%)
III-IV	417 (73.67%)	215 (79.63%)	56 (57.73%)	55 (74.32%)	79 (72.48%)
Unknown	14 (2.47%)	0	0	0	0
HPV status					
Positive	68 (12.01%)	60 (22.22%)	0 (%)	NA	NA
Negative	274 (48.41%)	209 (77.41%)	97 (100%)	NA	NA
Unknown	224 (39.58%)	1 (0.37%)	0	NA	NA
Radiotherapy					

HNSCC, head and neck squamous cell carcinoma; TCGA, The Cancer Genome Atlas; FHCRC, Fred Hutchinson Cancer Research Center; MDACC, MD Anderson Cancer Center; CSC, cancer stem cell; CSC-HR, CSC gene-associated high-risk; CSC-LR, CSC gene-associated low-risk; NA, not available

Characteristics	TCGA cohort (n = 566)	Leipzig cohort (n = 270)	FHCRC cohort (n = 97)	MDACC cohort (n = 74)	Greece cohort (n = 109)
Yes	304 (53.71%)	NA	NA	47 (63.51%)	54 (49.54%)
No	171 (30.21%)	NA	NA	26 (35.14%)	43 (39.45%)
Unknown	91 (16.08%)	NA	NA	1 (1.35%)	12 (11.01%)
Treatment					
Unimodal	188 (33.22%)	78 (28.89%)	43 (44.33%)	25 (33.78%)	43 (39.45%)
Multimodal	278 (49.12%)	189 (70.00%)	53 (54.64%)	48 (64.87%)	54 (49.54%)
Palliative	1 (0.17%)	3 (1.11%)	0	0	0
Unknown	99 (17.49%)	0	1 (1.03%)	1 (1.35%)	12 (11.01%)
CSC gene signature					
CSC-HR subgroup	285 (50.35%)	122 (45.19%)	38 (39.18%)	47 (63.51%)	57 (52.29%)
CSC-LR subgroup	281 (49.65%)	148 (54.81%)	59 (60.82%)	27 (36.49%)	52 (47.71%)

HNSCC, head and neck squamous cell carcinoma; TCGA, The Cancer Genome Atlas; FHCRC, Fred Hutchinson Cancer Research Center; MDACC, MD Anderson Cancer Center; CSC, cancer stem cell; CSC-HR, CSC gene-associated high-risk; CSC-LR, CSC gene-associated low-risk; NA, not available

CSC gene signature as an independent prognostic factor of HNSCC

To assess the independent prognostic factors of patients with HNSCC, we performed univariate and multivariate Cox proportional hazards models using CSC gene signature, patient demographics, social history, HPV status, and clinical staging of patients in TCGA and FHCRC HNSCC cohorts (n = 663). CSC gene signature (CSC-HR vs. CSC-LR subgroup), HPV status (negative vs. positive), and advanced clinical stage (stage III and IV vs. stage I and II) were independent prognostic factors of OS in patients with HNSCC (p = 0.0086, 0.0031, and 0.0035, respectively; Table S2).

Association of CSC gene signature with HPV status of HNSCC

We thought that if the additional survival analysis was performed individually according to the HPV status, it might be helpful to find appropriate indications to investigate CSC gene signatures to predict patient prognosis in HNSCC. Thus, we analyzed the prognosis of the CSC-HR and CSC-LR subgroups in

patients with HPV (+) and HPV (-) HNSCC from the five HNSCC cohorts (Fig. 3). There were no significant differences in 5-year OS rates between the CSC-HR and CSC-LR subgroups in patients with HPV (+) HNSCC (n = 128, p = 0.2; Fig. 3A). However, the CSC-HR subgroup showed significantly lower 5-year OS rates than the CSC-LR subgroup in patients with HPV (-) HNSCC (n = 578, p = 0.003; Fig. 3B).

Association of CSC gene signature with the result of radiotherapy

The expression of CSC markers is correlated with poor prognosis after radiotherapy in HNSCC [32, 33]. However, the clinical correlation between radiotherapy and genes encoding CSC markers has not been clearly studied. Thus, we analyzed the prognosis of the CSC-HR and CSC-LR subgroups in the five HNSCC cohorts that did and did not receive radiotherapy. The CSC-HR subgroup showed significantly lower 5-year OS rates than the CSC-LR subgroup among patients with HNSCC who received radiotherapy ($p < 0.0001$; Fig. 4A). But, the CSC-HR subgroup showed no significant differences in 5-year OS rates compared to the CSC-LR subgroup in patients with HNSCC who did not receive radiotherapy (Fig. 4B). Next, we compared patients who received radiotherapy and those who did not receive radiotherapy in the CSC-HR and CSC-LR subgroups, respectively. In the CSC-HR subgroup, patients who received radiotherapy showed no significant differences in 5-year OS rates compared to those who did not receive radiotherapy ($p = 0.1$; Fig. 4C). However, in the CSC-LR subgroup, patients who received radiotherapy showed significantly higher 5-year OS rates than those who did not receive radiotherapy ($p < 0.0001$; Fig. 4D). The interaction test for OS was performed to determine any correlation between radiotherapy and the CSC gene signature in HNSCC. The results revealed a significant correlation between radiotherapy and the CSC gene signature ($p < 0.0001$).

Pathway analysis

A total of 11 significant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were identified using DAVID (Table S3). Several pathways appeared to be related to the cancer or HNSCC pathways, including focal adhesion (p = 1.0E-5), small-cell lung cancer (p = 3.1E-4), ECM–receptor interaction (p = 4.6E-3), proteoglycans in cancer (p = 7.2E-3), PI3K-Akt signaling pathway (p = 1.0E-2), and pathways in cancer (p = 6.5E-2). Pathways associated with endothelial–mesenchymal transition signaling were also identified: regulation of the actin cytoskeleton (p = 8.5E-3) and leukocyte transendothelial migration (p = 9.9E-3).

Discussion

In this study, we developed and validated CSC gene signatures in five independent HNSCC cohorts. We observed that patients in the CSC-HR subgroup had a worse prognosis than those in the CSC-LR subgroup. Similar results were observed in the two subgroups of patients with HPV (-) HNSCC. Furthermore, the CSC gene signature could accurately predict outcomes in patients receiving radiotherapy. Thus, the CSC gene signature could identify patients with HNSCC who do not respond to radiotherapy and need intensified or personalized treatment.

Cancer cells within individual tumor masses often represent distinct phenotypic states that differ in their functional attributes [7]. Within this tumor heterogeneity, CSCs are considered essential for tumor initiation, maintenance, recurrence, and metastasis. To date, the identification of CSCs has been mainly based on CSC surface markers. However, genes encoding these CSC biomarkers to predict the prognosis of patients with HNSCC have not been studied clearly. Thus, we focused on the association between CSC biomarker genes and the prognosis of patients with HNSCC.

We thought it was difficult to predict the prognosis of HNSCC by considering all CSC biomarker genes, because each CSC biomarker gene had different effects on the prognosis, and the proportion of the expression of each CSC biomarker gene is heterogeneous depending on each patient. Thus, we decided to select CSC biomarker genes satisfying the following criteria: (a) whose corresponding biomarkers' high expression showed clinical significance in more than 2 studies in the last 10 years, and (b) whose high expression of each gene was significantly associated with prognosis in patients with HNSCC. Finally, we selected four CSC biomarker genes, CD44, MET, ALDH1A1, and BMI1. We then comprehensively analyzed five independent public cohorts considering these genes.

CD44 is a transmembrane glycoprotein that is the major receptor for hyaluronan [9]. CD44 is a commonly used CSC marker and is associated with prognosis in various human tumors, including HNSCC [34]. ALDH1 (aldehyde dehydrogenase 1) and BMI-1 (B-lymphoma Moloney murine leukemia virus insertion region-1) are two of the most studied CSC markers in HNSCC [35]. ALDH1 is an important stem cell marker for normal and cancer cells [36]. ALDH1 regulates cellular function by detoxifying various aldehydes and retinoid signaling. BMI-1 is important for the self-renewal ability of stem cells and is related to epithelial–mesenchymal transition [37]. The expression of c-MET (mesenchymal-to-epithelial transition factor) was found to be a CSC marker positively correlated with that of CD44 in HNSCC clinical databases [38]. From these results, the CSC biomarker genes selected in this study might also play a significant role in the prognosis of HNSCC.

Patients with HPV (-) HNSCC showed a worse prognosis in terms of OS and RFS than those with HPV (+) HNSCC [39]. However, each patient with HPV (-) HNSCC has a different prognosis because of various risk factors. We confirmed that the CSC gene signature was an independent prognostic factor for HNSCC regardless of HPV status. In addition, the CSC-HR subgroup showed a significantly worse prognosis than the CSC-LR subgroup in patients with HPV (-) HNSCC. Next, we investigated whether CSC gene signatures influenced the prognosis of patients with HPV (+) HNSCC. In these patients, the 5-year OS rates tended to be lower in the CSC-HR subgroup than in the CSC-LR subgroup, but the differences were not significant. This might be because of the relatively small size of the HPV (+) HNSCC cohorts (n = 128). Further studies in larger HPV (+) HNSCC cohorts are needed to confirm the association between CSC gene signatures and the prognosis of patients with HPV (+) HNSCC.

CSCs can regulate their proliferative and self-renewal capacity and, thus, are involved in metastasis, cancer development, and resistance to radiotherapy [40]. However, the association between various CSC biomarker genes and the response to radiotherapy in HNSCC has not been studied. Only the mRNA

expression of CD44 has been shown to be a significant predictor of local recurrence after radiotherapy in early-stage laryngeal cancer [32]. Thus, we hypothesized that the overexpression of a specific mRNA of CSC biomarker genes in HNSCC might be correlated with the response to radiotherapy. However, each patient heterogeneously expresses various CSC biomarker genes and, thus, might respond heterogeneously to radiotherapy. Our results showed that compared to the CSC-HR subgroup, the CSC-LR subgroup benefited significantly from radiotherapy. These results indicate that CSC gene signature might help to program a radiotherapy schedule if further research is conducted on the response to various doses of irradiation in CSC-HR and CSC-LR HNSCC cell lines.

A limitation of our study is that we analyzed CSC gene signatures using five different public HNSCC cohorts. Thus, there was a difference in the essential information included in each cohort. In particular, HPV status was missing in the two cohorts. In addition, detailed treatment modality methods or doses, such as postoperative radiotherapy, concurrent chemoradiotherapy, and induction chemoradiotherapy with operation, were not included in each cohort. To compensate for the missing information, we conducted additional analysis on the CSC gene signature and found that the CSC gene signature was associated with the prognosis in patients with HPV (-) HNSCC and the response to radiotherapy in HNSCC.

To the best of our knowledge, this is the first study to assess the prognosis of patients with HNSCC considering various CSC biomarker genes. Each CSC biomarker gene influenced the prognosis of patients with HNSCC, but the proportions of these genes were highly heterogeneous according to each patient. Thus, we clarified that the gene signatures of the four reference CSC biomarker genes—CD44, MET, ALDH1A1, and BMI1—were significantly related to the prognosis of patients with HNSCC. In addition, the Cox proportional hazards model showed that the CSC gene signature was an independent prognostic factor that influenced the OS of patients with HNSCC.

Conclusion

We developed CSC gene signatures that could predict the prognosis of patients with HNSCC. In addition, CSC gene signature was associated with the prognosis of patients with HNSCC who received radiotherapy or those with HPV (-) HNSCC. Therefore, our data provided evidence that CSC gene signatures may help design personalized treatments for patients with HNSCC following detailed classification.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets supporting the conclusions of this article are available in the UCSC Cancer Genomics Browser (http://xenabrowser.net/datapages/) (TCGA cohort, n=566) and the National Center for Biotechnology Information Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) (Leipzig cohort, GSE65858, n=270; FHCRC cohort, GSE41613, n=97; MDACC cohort, GSE42743, n=74; Greece cohort, GSE27020, n = 109).

Competing interests

The authors declare no conflicts interest.

Funding

None

Authors' contributions

SIK, SRW, YGE designed the study and wrote the manuscript text.

SIK, SRW, JKN, MKL, MK and SGK analyzed the results.

SIK, SRW, YCL, JWL, MK, and SGK designed the figures and tables.

YCL, JWL, and YGE reviewed and edited the paper.

All authors have read and approved the manuscript writing

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1A5A2019413) and by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI20C1205).

Authors' information (optional)

None

References

1. Kamangar F, Dores GM, Anderson WF: **Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the**

world. J Clin Oncol 2006, **24**(14):2137–2150.

2. Argiris A, Karamouzis MV, Raben D, Ferris RL: **Head and neck cancer**. Lancet 2008, **371**(9625):1695–1709.
3. Gupta S, Kong W, Peng Y, Miao Q, Mackillop WJ: **Temporal trends in the incidence and survival of cancers of the upper aerodigestive tract in Ontario and the United States**. Int J Cancer 2009, **125**(9):2159–2165.
4. Lohaus F, Linge A, Tinhofer I, Budach V, Gkika E, Stuschke M, Balermpas P, Rödel C, Avlar M, Grosu AL *et al*: **HPV16 DNA status is a strong prognosticator of loco-regional control after postoperative radiochemotherapy of locally advanced oropharyngeal carcinoma: results from a multicentre explorative study of the German Cancer Consortium Radiation Oncology Group (DKTK-ROG)**. Radiother Oncol 2014, **113**(3):317–323.
5. Linge A, Löck S, Gudziol V, Nowak A, Lohaus F, von Neubeck C, Jütz M, Abdollahi A, Debus J, Tinhofer I *et al*: **Low Cancer Stem Cell Marker Expression and Low Hypoxia Identify Good Prognosis Subgroups in HPV(-) HNSCC after Postoperative Radiochemotherapy: A Multicenter Study of the DKTK-ROG**. Clin Cancer Res 2016, **22**(11):2639–2649.
6. Rosen JM, Jordan CT: **The increasing complexity of the cancer stem cell paradigm**. Science 2009, **324**(5935):1670–1673.
7. Plaks V, Kong N, Werb Z: **The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells?** Cell Stem Cell 2015, **16**(3):225–238.
8. Ortiz RC, Lopes NM, Amôr NG, Ponce JB, Schmerling CK, Lara VS, Moyses RA, Rodini CO: **CD44 and ALDH1 immunoexpression as prognostic indicators of invasion and metastasis in oral squamous cell carcinoma**. J Oral Pathol Med 2018, **47**(8):740–747.
9. Huang CF, Xu XR, Wu TF, Sun ZJ, Zhang WF: **Correlation of ALDH1, CD44, OCT4 and SOX2 in tongue squamous cell carcinoma and their association with disease progression and prognosis**. J Oral Pathol Med 2014, **43**(7):492–498.
10. Xiao M, Liu L, Zhang S, Yang X, Wang Y: **Cancer stem cell biomarkers for head and neck squamous cell carcinoma: A bioinformatic analysis**. Oncol Rep 2018, **40**(6):3843–3851.
11. Picon H, Guddati AK: **Cancer stem cells in head and neck cancer**. Am J Stem Cells 2021, **10**(3):28–35.
12. Gupta PB, Chaffer CL, Weinberg RA: **Cancer stem cells: mirage or reality?** Nat Med 2009, **15**(9):1010–1012.
13. Wang T, Shigdar S, Gantier MP, Hou Y, Wang L, Li Y, Shamaileh HA, Yin W, Zhou SF, Zhao X *et al*: **Cancer stem cell targeted therapy: progress amid controversies**. Oncotarget 2015, **6**(42):44191–44206.
14. Kim SI, Woo SR, Noh JK, Lee MK, Lee YC, Lee JW, Ko SG, Eun YG: **Clinical significance of FAT1 gene mutation and mRNA expression in patients with head and neck squamous cell carcinoma**. Molecular oncology 2021.

15. Wichmann G, Rosolowski M, Krohn K, Kreuz M, Boehm A, Reiche A, Scharrer U, Halama D, Bertolini J, Bauer U *et al*: **The role of HPV RNA transcription, immune response-related gene expression and disruptive TP53 mutations in diagnostic and prognostic profiling of head and neck cancer**. Int J Cancer 2015, **137**(12):2846–2857.
16. Lohavanichbutr P, Méndez E, Holsinger FC, Rue TC, Zhang Y, Houck J, Upton MP, Futran N, Schwartz SM, Wang P *et al*: **A 13-gene signature prognostic of HPV-negative OSCC: discovery and external validation**. Clin Cancer Res 2013, **19**(5):1197–1203.
17. Fountzilas E, Kotoula V, Angouridakis N, Karasmanis I, Wirtz RM, Eleftheraki AG, Veltrup E, Markou K, Nikolaou A, Pectasides D *et al*: **Identification and validation of a multigene predictor of recurrence in primary laryngeal cancer**. PLoS One 2013, **8**(8):e70429.
18. Budczies J, Klauschen F, Sinn BV, Győrffy B, Schmitt WD, Darb-Esfahani S, Denkert C: **Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization**. PLoS One 2012, **7**(12):e51862.
19. Li J, Zhou J, Zhang J, Xiao Z, Wang W, Chen H, Lin L, Yang Q: **DNA repair genes are associated with tumor tissue differentiation and immune environment in lung adenocarcinoma: a bioinformatics analysis based on big data**. J Thorac Dis 2021, **13**(7):4464–4475.
20. Corona RI, Sudarshan S, Aluru S, Guo JT: **An SVM-based method for assessment of transcription factor-DNA complex models**. BMC Bioinformatics 2018, **19**(Suppl 20):506.
21. Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y: **Analysis of gene expression data using BRB-ArrayTools**. Cancer Inform 2007, **3**:11–17.
22. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. Nat Protoc 2009, **4**(1):44–57.
23. Jakob M, Sharaf K, Schirmer M, Leu M, Küffer S, Bertlich M, Ihler F, Haubner F, Canis M, Kitz J: **Role of cancer stem cell markers ALDH1, BCL11B, BMI-1, and CD44 in the prognosis of advanced HNSCC**. Strahlenther Onkol 2021, **197**(3):231–245.
24. Mannelli G, Magnelli L, Deganello A, Busoni M, Meccariello G, Parrinello G, Gallo O: **Detection of putative stem cell markers, CD44/CD133, in primary and lymph node metastases in head and neck squamous cell carcinomas. A preliminary immunohistochemical and in vitro study**. Clin Otolaryngol 2015, **40**(4):312–320.
25. Baschnagel AM, Williams L, Hanna A, Chen PY, Krauss DJ, Pruetz BL, Akervall J, Wilson GD: **c-Met expression is a marker of poor prognosis in patients with locally advanced head and neck squamous cell carcinoma treated with chemoradiation**. Int J Radiat Oncol Biol Phys 2014, **88**(3):701–707.
26. Rao RS, Raju KL, Augustine D, Patil S: **Prognostic Significance of ALDH1, Bmi1, and OCT4 Expression in Oral Epithelial Dysplasia and Oral Squamous Cell Carcinoma**. Cancer Control 2020, **27**(1):1073274820904959.
27. Michifuri Y, Hirohashi Y, Torigoe T, Miyazaki A, Kobayashi J, Sasaki T, Fujino J, Asanuma H, Tamura Y, Nakamori K *et al*: **High expression of ALDH1 and SOX2 diffuse staining pattern of oral squamous cell carcinomas correlates to lymph node metastasis**. Pathol Int 2012, **62**(10):684–689.

28. Tamatani T, Takamaru N, Ohe G, Akita K, Nakagawa T, Miyamoto Y: **Expression of CD44, CD44v9, ABCG2, CD24, Bmi-1 and ALDH1 in stage I and II oral squamous cell carcinoma and their association with clinicopathological factors**. Oncol Lett 2018, **16**(1):1133–1140.
29. Hu Q, Wu T, Chen X, Li H, Du Z, Hao Y, Peng J, Tai S, Song M, Cheng B: **The poor outcome of second primary oral squamous cell carcinoma is attributed to Bmi1 upregulation**. Cancer Med 2018, **7**(4):1056–1069.
30. Canis M, Lechner A, Mack B, Zengel P, Laubender RP, Koehler U, Heissmeyer V, Gires O: **CD133 is a predictor of poor survival in head and neck squamous cell carcinomas**. Cancer Biomark 2012, **12**(2):97–105.
31. Habu N, Imanishi Y, Kameyama K, Shimoda M, Tokumaru Y, Sakamoto K, Fujii R, Shigetomi S, Otsuka K, Sato Y *et al*: **Expression of Oct3/4 and Nanog in the head and neck squamous carcinoma cells and its clinical implications for delayed neck metastasis in stage I/II oral tongue squamous cell carcinoma**. *BMC Cancer* 2015, **15**:730.
32. de Jong MC, Pramana J, van der Wal JE, Lacko M, Peutz-Kootstra CJ, de Jong JM, Takes RP, Kaanders JH, van der Laan BF, Wachters J *et al*: **CD44 expression predicts local recurrence after radiotherapy in larynx cancer**. Clin Cancer Res 2010, **16**(21):5329–5338.
33. Koukourakis MI, Giatromanolaki A, Tsakmaki V, Danielidis V, Sivridis E: **Cancer stem cell phenotype relates to radio-chemotherapy outcome in locally advanced squamous cell head-neck cancer**. Br J Cancer 2012, **106**(5):846–853.
34. Orian-Rousseau V: **CD44 Acts as a Signaling Platform Controlling Tumor Progression and Metastasis**. Front Immunol 2015, **6**:154.
35. Sharaf K, Lechner A, Haider SP, Wiebringhaus R, Walz C, Kranz G, Canis M, Haubner F, Gires O, Baumeister P: **Discrimination of Cancer Stem Cell Markers ALDH1A1, BCL11B, BMI-1, and CD44 in Different Tissues of HNSCC Patients**. Curr Oncol 2021, **28**(4):2763–2774.
36. Vassalli G: **Aldehyde Dehydrogenases: Not Just Markers, but Functional Regulators of Stem Cells**. Stem Cells Int 2019, **2019**:3904645.
37. Curtarelli RB, Gonçalves JM, Dos Santos LGP, Savi MG, Nör JE, Mezzomo LAM, Rodríguez Cordeiro MM: **Expression of Cancer Stem Cell Biomarkers in Human Head and Neck Carcinomas: a Systematic Review**. Stem Cell Rev Rep 2018, **14**(6):769–784.
38. Lee JC, Wu ATH, Chen JH, Huang WY, Lawal B, Mokgautsi N, Huang HS, Ho CL: **HNC0014, a Multi-Targeted Small-Molecule, Inhibits Head and Neck Squamous Cell Carcinoma by Suppressing c-Met/STAT3/CD44/PD-L1 Oncoimmune Signature and Eliciting Antitumor Immune Responses**. Cancers (Basel) 2020, **12**(12).
39. Gupta AK, Lee JH, Wilke WW, Quon H, Smith G, Maity A, Buatti JM, Spitz DR: **Radiation response in two HPV-infected head-and-neck cancer cell lines in comparison to a non-HPV-infected cell line and relationship to signaling through AKT**. Int J Radiat Oncol Biol Phys 2009, **74**(3):928–933.
40. Senobari Z, Karimi G, Jamialahmadi K: **Ellagitannins, promising pharmacological agents for the treatment of cancer stem cells**. Phytother Res 2021.

Figures

Figure 1

Stratification of patients with HNSCC in TCGA cohort according to the 81 CSC gene signatures.

(a) Patients in TCGA HNSCC cohort were classified into CSC-HR (n=285) and CSC-LR (n=281) subgroups by hierarchical clustering. (b, c) The 5-year OS and RFS rates of each group were determined using Kaplan–Meier plots. The CSC-HR subgroup showed significantly lower 5-year OS and RFS rates than the CSC-LR subgroup (p=0.04 and 0.02, respectively). *p<0.05

Figure 2

Validation of the 81 CSC gene signatures in the four independent HNSCC cohorts.

Each cohort was stratified to CSC-HR and CSC-LR subgroups by using the 81 CSC gene signatures. Predicted outcomes in four cohorts were depicted using Kaplan–Meier plots. (a-c) The 5-year OS rates of each subgroup were determined in the Leipzig (n=270), FHCRC (n=97), and MDACC (n=109) cohorts (p=0.06, <0.0001, and =0.02, respectively). (d) Furthermore, the 5-year RFS rates of each subgroup was determined in the Greece cohort (n=109; p=0.009). *p<0.05

Fig. 3

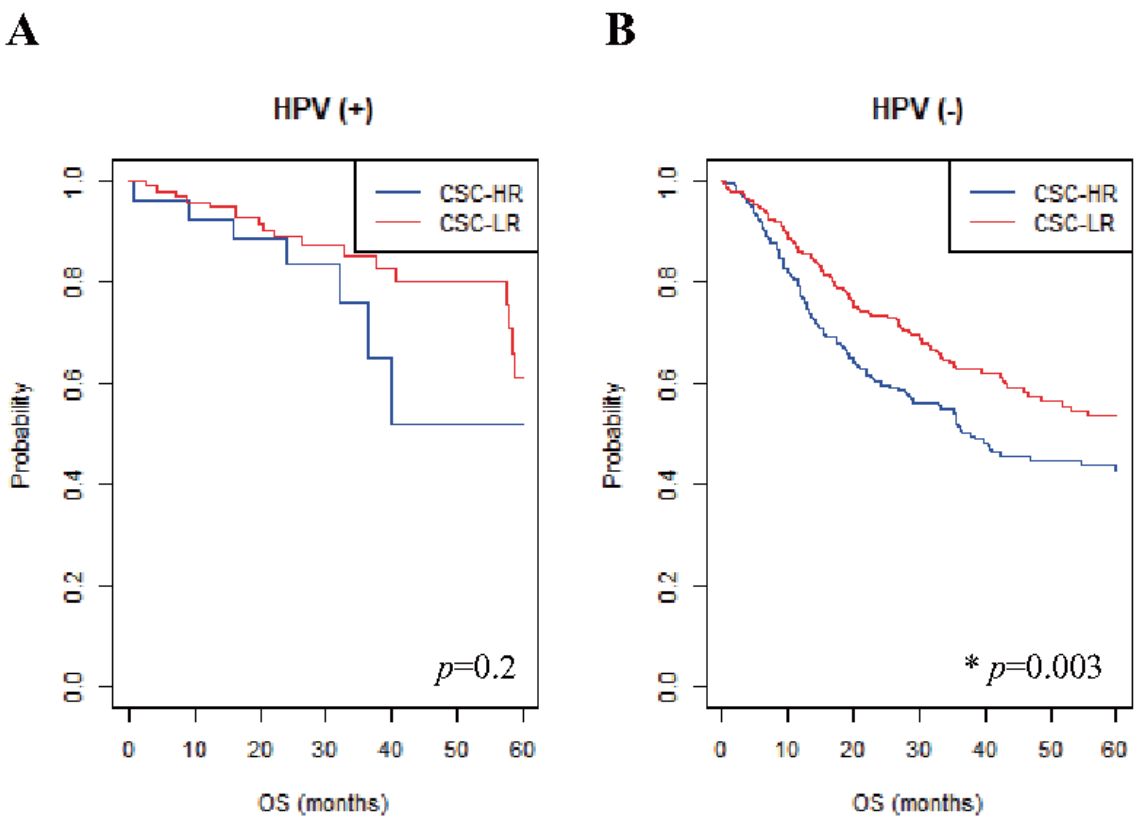

Figure 3

Association of CSC gene signature with HPV status in the five independent HNSCC cohorts.

(a, b) The 5-year OS rates of CSC-HR and CSC-LR subgroups in patients with HPV (+) and HPV (-) HNSCC are depicted using Kaplan–Meier plots (n=128 and 578; p=0.2 and 0.003, respectively). *p<0.05

Fig. 4

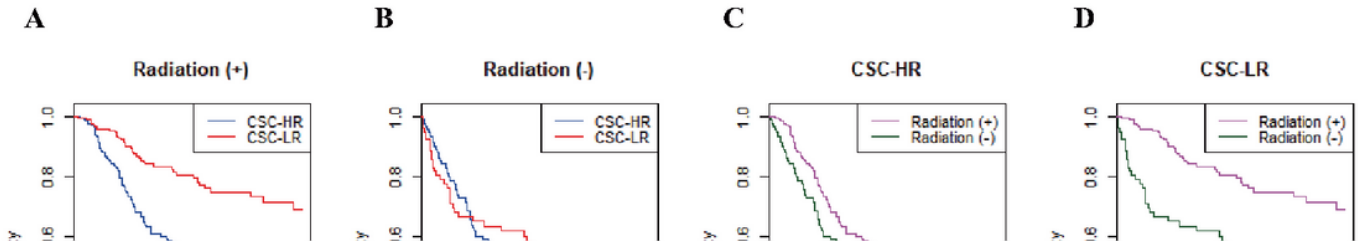

Figure 4

Association of CSC gene signature with radiotherapy in the five independent HNSCC cohorts.

(a, b) The 5-year OS rates of CSC-HR and CSC-LR subgroups in patients with HNSCC who did and did not receive radiotherapy (n=348 and 196, respectively) and (c, d) those of patients with HNSCC according to radiotherapy in the CSC-HR (n=293) and CSC-LR (n=251) subgroups are depicted using Kaplan–Meier plots. Patients in the CSC-LR subgroup benefited significantly from radiotherapy ($p<0.0001$ and <0.0001, respectively). *$p<0.05$

Supplementary Files

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

- CSCtablesupplementary13.docx
- supplefig1.pdf
- supplefig2.pdf