Oral Oncology 51 (2015) 355–362

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

An eleven gene molecular signature for extra-capsular spread in oral squamous cell carcinoma serves as a prognosticator of outcome in patients without nodal metastases

Weining Wang [a], Weng Khong Lim [b], Hui Sun Leong [c], Fui Teen Chong [c], Tony K.H. Lim [d], Daniel S.W. Tan [b,e], Bin Tean Teh [c], N. Gopalakrishna Iyer [a,b,c,*]

[a] Department of Surgical Oncology, National Cancer Centre, 11 Hospital Drive, Singapore 169610, Singapore
[b] Cancer and Stem Cell Biology, Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857, Singapore
[c] Cancer Therapeutics Research Laboratory, National Cancer Centre, 11 Hospital Drive, Singapore 169610, Singapore
[d] Department of Pathology, Singapore General Hospital, Outram Road, Singapore 169610, Singapore
[e] Department of Medical Oncology, National Cancer Centre, 11 Hospital Drive, Singapore 169610, Singapore

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form 1 December 2014
Accepted 13 December 2014
Available online 2 January 2015

Keywords:
Extracapsular spread
Gene molecular signature
Oral squamous cell carcinoma
Prognostic factor

SUMMARY

Objectives: Extracapsular spread (ECS) is an important prognostic factor for oral squamous cell carcinoma (OSCC) and is used to guide management. In this study, we aimed to identify an expression profile signature for ECS in node-positive OSCC using data derived from two different sources: a cohort of OSCC patients from our institution (National Cancer Centre Singapore) and The Cancer Genome Atlas (TCGA) head and neck squamous cell carcinoma (HNSCC) cohort. We also sought to determine if this signature could serve as a prognostic factor in node negative cancers.

Materials and methods: Patients with a histological diagnosis of OSCC were identified from an institutional database and fresh tumor samples were retrieved. RNA was extracted and gene expression profiling was performed using the Affymetrix GeneChip Human Genome U133 Plus 2.0 microarray platform. RNA sequence data and corresponding clinical data for the TCGA HNSCC cohort were downloaded from the TCGA Data Portal. All data analyses were conducted using R package and SPSS.

Results: We identified an 11 gene signature (GGH, MTFR1, CDKN3, PSRC1, SMIM3, CA9, IRX4, CPA3, ZSCAN16, CBX7 and ZFP3) which was robust in segregating tumors by ECS status. In node negative patients, patients harboring this ECS signature had a significantly worse overall survival ($p = 0.04$).

Conclusions: An eleven gene signature for ECS was derived. Our results also suggest that this signature is prognostic in a separate subset of patients with no nodal metastasis Further validation of this signature on other datasets and immunohistochemical studies are required to establish utility of this signature in stratifying early stage OSCC patients.

Introduction

Oral squamous cell carcinoma (OSCC) poses a significant global public health threat, with less than 50% of patients surviving beyond five years [1]. Most of these originate from the anterior tongue, although other sites may be affected, such as the alveolus, floor of mouth, retromolar trigone and hard palate. Despite advances in the detection and treatment of these cancers, survival rates have not changed in the last two decades [1,2]. Given that the majority of OSCC cases occur in developing countries, a public health approach to prevention and early detection remains key to improving outcomes. However, there is pressing need to recognize that oral cancers are a heterogeneous group of cancers that may require individualized approaches to treatment.

One unique feature in the management of OSCC is the role of primary surgery as the main treatment modality. Several studies have indicated that treatment with primary surgery (and adjuvant radio- or chemoradio-therapy in selected high-risk cases) results in improved survival outcomes, compared with primary radiotherapy as a treatment modality [3,4]. Despite this approach, a sizeable proportion of patients relapse and develop disease recurrence either locally at the primary site, or present with metastasis at

* Corresponding author at: Department of Surgical Oncology, National Cancer Centre Singapore, 11 Hospital Drive, Singapore 169610, Singapore. Tel.: +65 64368294; fax: +65 62257559.
E-mail address: gopaliyer@yahoo.com (N.G. Iyer).

http://dx.doi.org/10.1016/j.oraloncology.2014.12.012
1368-8375/

regional lymph nodes or distant sites [5]. The identification of patients at high risk of relapsing is thus a key research area in OSCC. Current strategies for risk stratification in OSCC have been based on pathologic features associated with poor outcome such as tumor stage, nodal metastasis, perineural or lymphovascular invasion and extracapsular spread (ECS) [6–9]. Several studies have reported ECS as one of the most important risk factors for treatment failure in OSCC, and ECS has been shown to be an important criteria in selecting patients for the addition of concurrent chemotherapy to post-operative radiotherapy [10,11]. ECS remains a significant risk factor for recurrence, distant metastases and death despite current treatment protocols. Initially attributed to a purely mechanical extension and invasion seen in larger tumor deposits in lymph nodes, there is increasing evidence that tumors that develop ECS and consequently, soft tissue extension, are biologically distinct, and have increased tendency to soft tissue invasion and distant metastases. This is seen in a number of studies where the ECS does not correlate in a linear manner with size of nodal metastases, and that ECS is even seen in patients with microscopic lymph node metastases [12–14].

In this study, we aimed to identify an expression profile signature for ECS in node-positive OSCC using data derived from two different sources: a cohort of OSCC patients from our institution (National Cancer Centre Singapore) and The Cancer Genome Atlas (TCGA) head and neck squamous cell carcinoma (HNSCC) cohort. We then sought to determine if this signature could serve as a prognostic factor in node negative cancers.

Materials and methods

Tumor and patients

Patients included in the local cohort were identified from an institutional database of consecutive patients treated at the National Cancer Centre Singapore (NCCS) between January 1998 and March 2009. Included patients were confirmed to have a histological diagnosis of squamous cell carcinoma involving the oral cavity, with complete clinico–pathologic data and fresh frozen tumor samples available. Fresh tumor samples were retrieved from Singhealth Tissue Repository. Tumor content for each sample was assessed by microscopic examination of hemotoxylin and eosin (H&E) stained sections of the tissue by a board-certified pathologist (TKL), and confirmed to be at least 50%. The study was approved by the Singhealth Centralized Institutional Review Board (CIRB 2008/467/B).

Tissue preparation and RNA extraction

RNA was extracted from a 3 mm × 3 mm section of fresh frozen tumor using the RNeasy Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. RNA concentration and purity were quantified and assessed using the Nanodrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE). Our inclusion criteria was sufficient RNA quality (A260/A280 ratio >1.60), as well as complete clinico–pathologic and outcome data. A total of 57 samples met the criteria and were retained for the study (Table 1).

Gene expression profiling data

Gene expression profiling was performed using the Affymetrix GeneChip Human Genome U133 Plus 2.0 microarray platform according to manufacturer's instructions (Affymetrix, Santa Clara, CA). Briefly, extracted and purified RNA were converted to complementary DNA (cDNA). An in vitro transcription step was then performed to generate complementary RNA (cRNA) that is hybridized to the microarray. Unhybridized cRNA were then washed off while the bound cRNA were fluorescently labelled. The microarrays were then scanned using an Affymetrix GeneChip Scanner 3000 confocal laser scanner.

The resulting CEL files were then loaded into the R statistical environment (http://www.r-project.org) using the *simpleaffy* package [15]. Normalization of the data was done using the Robust Multi-chip Average (RMA) method [16]. A custom chip definition file (CDF) from the BrainArray project was then used to summarize probe sets to genes to facilitate analysis [17].

TCGA level 2 RNA-seq data for HNSCC samples were downloaded from the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga, accessed on April 18th 2013) along with patient clinical annotation. Raw gene-level counts were loaded into R, and then processed using the *voom* (variance modelling at observational level) tool in the *limma* package with ECS status as covariates. The function takes raw counts, performs quantile normalization to account for between-sample variation, and then generates cpm (counts per-million) values that is comparable across samples of varying library sizes.

Data analysis

Differential expression analysis was performed in R using the *limma* package [18]. An empirical Bayes moderated *t*-test was used to identify differentially expressed genes. A fold-change threshold of 1.5× was used and genes with p-values < 0.05 were considered to be significantly differentially expressed. Heatmaps were generated using the *gplots* R package, where hierarchical clustering of samples and genes was performed using Pearson correlation as a distance metric and Ward's clustering method. K-means clustering of samples was done using the *kmeans* function in R with two centers and 1000 random iterations. Outcome analyses were performed using the Kaplan–Meier method and the log-rank test was used to compare the difference between treatment arms. The latter analyses were performed using using PASW statistics version 18.0 (IBM, New York, NY), and a p value of less than 0.05 was deemed significant.

Results

In order to identify a set of genes consistently differentially expressed between node-positive patients with ECS and those without, we used two independent datasets generated using orthogonal technologies. First, a subset of 34 node-positive OSCC patients in the TCGA HNSCC with available ECS status was used. In this dataset, a total of 578 genes were found to be differentially expressed between node-positive OSCC patients with ECS and those without (fold-change >1.5× and p-value < 0.05, Fig. 1). Second, a similar comparison was repeated on a cohort of 20 node-positive OSCC patients that was treated at NCCS. In this cohort, a total of 338 genes were differentially expressed (fold-change >1.5×, p-value < 0.05, Fig. 2) between ECS and non-ECS patients. We then intersected up- and downregulated genes from these two datasets to obtain a set of genes consistently dysregulated in association with ECS both TCGA and NCCS cohorts; an exercise that yielded 11 genes. We confirmed that this 11-gene signature is robust in segregating samples by ECS status by using it to perform K-means clustering ($k = 2$) of the NCCS dataset (Fig. 3, two-tailed Fisher's exact test p-value = 0.003).Interestingly, of the three cases that were 'misclassified' with having the ECS signature, a review of pathology showed that two demonstrated lymphovascular or perineural invasion, and both developed recurrence within 15 months. The 11 gene set included the following genes: GGH, MTFR1, CDKN3, PSRC1, SMIM3, CA9, IRX4, CPA3, ZSCAN16, CBX7 and ZFP3, and their known or putative functions are described in Table 2.