Expression of defensin-associated gene can be correlated with lymph node metastasis of early stage tongue cancer

Doh Young Lee¹*, J. Hun Hah²*, Woo-Jin Jeong³, Eun-Jae Chung⁴, Tack-Kyun Kwon¹, Soon-Hyun Ahn⁴, Myung-Whun Sung⁴, Seong Keun Kwon⁴,⁵,⁶

¹Department of Otorhinolaryngology Head and Neck Surgery, Seoul Nation University Boramae Medical Center;

²Department of Otorhinolaryngology Head and Neck Surgery, THANQ Seoul Thyroid-Head & Neck Surgery Center;

³Department of Otorhinolaryngology Head and Neck Surgery, Seoul National University Bundang Hospital;

⁴Department of Otorhinolaryngology Head and Neck Surgery Seoul National University Hospital.

⁵Sensory Organ Research Institute, Seoul National University Medical Research Center

⁶Cancer Research Institute, Seoul National University, Seoul 03080, Republic of Korea

*These two authors (DY Lee and JH Hah) equally contributed to this work.

Running title: Defensin in tongue cancer metastasis

**Corresponding author**

Seong Keun Kwon, M.D., Ph.D.

Professor

Department of Otorhinolaryngology-Head and Neck Surgery

Seoul National University Hospital 100 Daehak-Ro Jongno-Gu, Seoul, Korea

Tel: 82-2-2072-0215
Fax: 82-2-745-2387
E-mail: otolarynx@hanmail.net

**Conflict of interest:** The authors declare that they have no potential conflicts of interest.

**Financial disclosure:** This work has supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018R1D1A1B07047534). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2017R1A2B4011229). This study was supported by a clinical research grant provided from Seoul National University Boramae Medical Center.

**ORCID ID**
Doh Young Lee: 0000-0003-1590-8559
J. Hun Hah: 0000-0002-5110-012X
Woo-Jin Jeong: 0000-0007-8577-6964
Eun-Jae Chung: 0000-0003-3918-7717
Tack-Kyun Kwon: 0000-0001-8250-914X
Soon-Hyun Ahn: 0000-0002-0759-6850
Myung-Whun Sung: 0000-0003-3513-9996
Seong Keun Kwon: 0000-0001-9218-7666

**Conceptualization:** DYL, JHH, SKK.

**Data curation:** DYL, JHH.
Formal analysis: DYL.

Methodology: JHH, WJJ, EJC, TKK, SHA, MWS.

Visualization: DYL, SKK.

Writing original draft: DYL, SKK. Writing review & editing: DYL, SKK.
HIGHLIGHTS

In this study, we aimed to analyze the profile of gene expression for selecting patients eligible for neck dissection of T1/2 and node negative tongue cancer.

We found that Defensin (DEFB4A, DEFB103B, DEFB4B) may be a novel biomarker for early regional metastasis in T1/2 tongue cancer.
ABSTRACT

BACKGROUND: We aimed to assess the genetic differences between positive and negative lymph node metastasis in early stage tongue cancer.

METHODS: A total of 35 cases of tongue cancer with RNAseq data were enrolled in this study. The gene expression profile of the following two groups was compared: 1) N0 group, T stage 1 or 2 with N0 stage; 2) N+ group, T stage 1 or 2 with N+ stage. Using the R and limma packages in the Bioconductor program, we extracted the differentially expressed genes (DEGs). Gene ontology and pathway enrichment analysis were performed using the DAVID online tool. Immune cell infiltration was analyzed using the CIBERSORT online program. Immunochemical staining of the cancer tissue was evaluated and the Cancer Genome Atlas (TCGA) data were analyzed to validate of the identified DEGs.

RESULTS: There were no significant differences in infiltration of the 22 types of immune cells. Among a total of 51 DEGs identified, 14 genes were significantly upregulated while 37 genes were significantly downregulated (p<0.01; fold-change >2). Pathway analysis revealed that arachidonic acid metabolism-related pathway, calcium, and muscle contraction pathway were significantly correlated. Among DEGs, following genes were the most significantly different between two groups; DEFB4A, SPRR2B, DEFB103B, SPRR2G, DEFB4B, and FAM25A. Via TCGA data, DEFB4A and DEFB103B were more highly expressed in the N0 group than in the N+ group, although the difference did not achieve statistical significance. Immunochemical staining of cancer tissue revealed the significantly higher expression of defensin in the N0 group.

CONCLUSIONS: Defensin (DEFB4A, DEFB103B, DEFB4B) may be a novel biomarker for early regional metastasis in T1/2 tongue cancer.
Keywords: tongue cancers, squamous cell carcinoma, metastasis, defensins
INTRODUCTION

In oral cavity cancer, 6% to 85% of cases are known to have regional metastasis [1]. As the extension of lymph node metastasis in neck are one of the most important prognostic factors, elective neck dissection, biopsy of sentinel lymph node, radiation therapy, and observation are possible treatment option for clinically lymph node negative oral cavity cancer. Among oral cavity cancers, tongue squamous cell carcinoma is the well-studied subsite [2]. Several clinicoradiologic and pathologic parameters are suggested and one of the most important parameters is the invasion depth of the primary tumor [3]. However, the limitation of using depth of invasion cutoff is the difficulty in obtaining this information preoperatively. To overcome this limitation, staged neck dissection or preoperative deep biopsy can be another option, while the clinical feasibility is low because of sampling error and lack of surgeon’s preference. T stage is another parameter for predicting regional metastasis in clinically node negative oral cancer [4]. It can be suggested that most of cases with tongue cancer higher than T2 should be the candidate for elective neck dissection.

Elective lymph node dissection has been a part of the treatment of patients with clinically node negative tongue squamous cell carcinoma. The main disadvantage of this procedure is that it may produce significant disability and scar; thus, ideally it should be avoided in patients with pathologically negative lymph nodes. Till recently, limited number of studies using genetic information of tongue cancer have evaluated the association between genetic profiles and early regional metastasis of tongue cancer. We aimed to analyze difference in gene expression profile between positive and negative lymph node metastasis in oral tongue cancer patients and to search genetic marker for selecting eligible patient for elective neck
dissection.
METHODS

This study was approved by the Institutional Review Boards (IRB) of *** hospital (IRB No. [Redacted]) and the *** Center (IRB No. [Redacted]). Written informed consent was obtained from all participants, and all of the methods were performed in accordance with the relevant guideline and regulation.

**Human tumor samples**

All samples were collected and preserved in the *** Hospital Human Biobank. Among the head and neck cancer patients who were diagnosed and surgically treated, a total of 49 fresh tongue squamous cell carcinoma samples were collected from Jan 2012 to Dec 2016. Among the samples, final pathologic T stage 1 and 2 nonrecurrent cases were selected for evaluation (n=35; Supplementary Table 1). All of the enrolled patients were clinically N0, which means that preoperative imaging studies (positron emission tomography, computed tomography, and magnetic resonance imaging) revealed that there was no suspicious metastatic lymph node in neck. Patients in N0 group underwent elective neck dissection, and final pathologic lymph node status were as shown in Supplementary Table 1. Age, sex, tumor size and thickness, nearest resection margin, and level and number of metastatic lymph nodes were collected for the patients from which the samples were collected. To analyze the differences in expression of the transcriptome in tumors with early regional metastases, 35 patients were categorized into 2 groups: (1) the N0 group (n=23) with no regional metastasis (pN0) and (2) the N+ group (n=12) with positive regional metastasis (pN1 and 2).
**mRNA sequencing**

Total RNA concentration was calculated by Quant-IT RiboGreen (Invitrogen). To determine the DV200 (% of RNA fragments >200 bp) value, samples are run on the TapeStation RNA screentape (Agilent). A total of 100 ng of total RNA was subjected to a sequencing library construction using a TruSeq RNA Access library prep kit (Illumina, San Diego, CA USA) according to the manufacturer's protocols. Briefly, the total RNA was firstly fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments are copied into first strand cDNA using SuperScript II reverse transcriptase (Invitrogen, #18064014) and random primers. This is followed by second strand cDNA synthesis using DNA Polymerase I, RNase H and dUTP. These cDNA fragments then go through an end repair process, the addition of a single ‘A’ base, and then ligation of the adapters. The products are then purified and enriched with PCR to create the cDNA library. All libraries were normalized and six libraries were pooled into a single hybridization/capture reaction. Pooled libraries were incubated with a cocktail of biotinylated oligos corresponding to coding regions of the genome. Targeted library molecules were captured via hybridized biotinylated oligo probes using streptavidin-conjugated beads. After two rounds of hybridization/capture reactions, the enriched library molecules were subjected to a second round of PCR amplification. The captured libraries were quantified using KAPA Library Quantification kits for Illumina Sequencing platforms according to the qPCR Quantification Protocol Guide (KAPA BIOSYSTEMS, #KK4854) and qualified using the TapeStation D1000 ScreenTape (Agilent Technologies, # 5067-5582). Indexed libraries were then submitted to an Illumina Hiseq2500 (Illumina, Inc., San Diego, CA, USA), and the paired-end (2×100 bp) sequencing.
was performed by the Macrogen Incorporated.

We preprocessed the raw reads from the sequencer to remove low quality and adapter sequence before analysis and aligned the processed reads to the Homo sapiens (hg19) using HISAT v2.0.5 [5]. HISAT utilizes two types of indexes for alignment (a global, whole-genome index and tens of thousands of small local indexes). These two types’ indexes are constructed using the same BWT (Burrows–Wheeler transform)/ a graph FM index (GFM) as Bowtie2. Because of its use of these efficient data structures and algorithms, HISAT generates spliced alignments several times faster than Bowtie and BWA widely used. The reference genome sequence of Homo sapiens (hg19) and annotation data were downloaded from the UCSC table browser (http://genome.uscs.edu). Transcript assembly and abundance estimation using StringTie [6,7]. After alignment, StringTie v1.3.3b was used to assemble aligned reads into transcripts and to estimate their abundance. It provides the relative abundance estimates as FPKM values (Fragments Per Kilobase of exon per Million fragments mapped) of transcript and gene expressed in each sample.

**Differentially expressed gene analysis and gene ontology (GO) enrichment / KEGG pathway analysis**

Differentially expressed genes (DEGs) between the N0 and N+ groups were identified by two-tailed Student’s t-tests using the Linear Models for Microarray Data (Limma) package in R. The alpha level was 0.4 and adjusted p-values were obtained. DEG was defined as the genes with fold change>2 with adjusted p-value<0.01.

The Database for Annotation, Visualization and Integration Discovery (DAVID) web-based
program was utilized for gene ontology enrichment analysis and functional annotation. DAVID network software (NIH, Bethesda, MD, USA) consist of the most of major public resources of bioinformatics data. Gene-related biological mechanisms with standardized gene terminology can be analyzed using DAVID. In addition, high-throughput gene functional analysis can be facilitated by the DAVID knowledge base. The set of DEGs and the associated biological pathway information for the gene sets were analyzed by the DAVID.

**Analysis of immune cell infiltration**

According to the LM22 signature file, which is the signature profiles for 22 distinct immune cell types, gene expression data were analyzed in Cibersort program [8].

**Validation using TCGA data and immunohistochemical stain**

Among the TCGA Illumina Hiseq Rnaseqv2 data and matched clinical data, oral tongue cancer patients with T stage 1 or 2 were included. A total of 70 cases were divided by status of lymph node metastasis: N0 (39 cases) and N+ (31 cases). Using the Student’s t-test, DEG expression was compared between two groups.

The paraffin sections (4μm) of tumor tissues were cut, deparaffinized, and treated with 0.03% H₂O₂ in methanol for 10 minutes to quench endogenous peroxidase activity. After washing in PBS, the sections were then incubated overnight at 4 °C with beta-defensin 3 antibody (1:10, Novus Biologicals, CO, USA). During the next day, the sections were treated with secondary antibody (Vector Laboratories) in PBS (1:400) for 2 hours at room temperature. Then, antigen-antibody complexes were detected using an avidin-biotin complex detection system.
(Vectastain ABC Kit, Vector Laboratories) and a DAB Substrate Kit (Vector Laboratories). After counterstaining with Mayer’s hematoxylin, the sections were examined using an Olympus BX51 microscope. The pictures were captured in Olympus DP72 and DP2-BSW (Olympus, Tokyo, Japan). Histological sections were analyzed using Image J Software (NIH, Bethesda, MD, USA), and semiquantitative analysis was performed by intensity of staining.
RESULTS

Clinicopathologic data

N0 and N+ groups showed no differences in age, sex, tumor size and thickness, or nearest resection margin (Table 1). Recurrence was higher in the N+ group, although it did not attain statistical significance.

Differentially expressed genes and difference in immune cell infiltration

Among a total of 51 DEGs, up-regulation was in 14 genes and down-regulation was in 37 genes (p<0.01; fold-change >2; Fig 1). Via heatmap analysis, the N0 and N+ groups showed moderate clustering of DEGs (Fig. 2). The most significant difference was found in the six following: DEFB4A, SPRR2B, DEFB103B, SPRR2G, DEFB4B, and FAM25A. There was no significant difference of the infiltration of the 22 different types of immune cells between the N0 and N+ groups (Fig. 3). Additionally, the ratio of M1 to M2 macrophages also showed no significant difference (p=0.661).

GO enrichment / KEGG pathway analysis and expression of DEFB in TCGA

Enriched DEGs are listed in Table 2. The highest enrichment score of 7.39 was associated ‘secreted’, ‘signal’, and ‘disulfide bond’. The second highest score of 4.83 was associated with ‘protease inhibitor’, ‘serine protease inhibitor’, and ‘negative regulation of endopeptidase activity’. DEFB, which had the most significantly different expression
between the two groups, was associated with ‘antibiotic’, ‘antimicrobial’, ‘beta defensing type’, defense response to bacterium’, and ‘defensin’. In the TCGA database, DEFB4A and DEFB103B were more highly expressed in the N0 group than in the N+ group, although this difference did not attain statistical significance (Fig. 4).

**Immunohistochemical staining**

Immunohistochemical staining showed significantly higher expression of defensin in the N0 group. (p<0.01) (Fig. 5)
Gene expression profile has been shown to be feasible for the determining prognosis and early detection of several cancers [9]. Previous studies have demonstrated that profile of gene expression can potential indicator for prediction of lymph node metastasis in oral cavity cancers [10]. Watanabe et al. reported that the expression profile of a panel of 19 genes allowed the prediction of lymphatic metastasis of oral tongue cancer [11]. The prediction of lymph node metastasis is very important in tongue cancer with early stage because an accurate prediction may prevent unnecessary neck dissection. Although tumor thickness is a very useful and strongly reliable parameter for cervical lymph node metastasis, correct measurement of pathologic thickness is near impossible in a preoperative setting. Gene expression can be good alternative for lymph node metastasis in tongue cancer with early stage. Van Hooff et al. reported that combination of gene expression profile and current clinical parameters could reduce the rate of pathologically node negative in elective neck dissection from 28% to 11% in early stage oral squamous cell carcinoma [12]. They concluded that a reduced rate of undetected nodal metastasis should be sufficient to encourage clinicians not to perform unnecessary neck treatment in patients with node negative.

In our study, reduced expression of defensin related genes (DEFB4A, DEFB103B, and DEFB4B) were associated regional metastasis of neck in tongue cancer with early stage. The innate immune system has a critical role for defending host against pathogens in addition to the adaptive immunity to bacteria, fungi, and viruses is developed. In particular, mucosal cells generate various kinds of antimicrobial peptides which prohibit the growth and invasion
of many kinds of pathogens. Among those, defensins are cationic antimicrobial peptides [13-15], and there are 6 and 28 human α-defensins and β-defensins, respectively. β-defensins are usually located in the skin, airways, uterus and oral cavity [16,17]. Human β-defensins is known to connect and interact adaptive immunity and innate immunity by biding to CCR6 and increasing the chemotactic activity for memory T cells and immature dendritic cells [18].

Uraki et al. reported that migration of colon cancer cells can be inhibited by human β-defensin-3 [19]. They found that human β-defensin-3 is expressed in tumor-infiltrating monocytes, and not expressed in colon cancer cells. In addition, migration of various colon cancer cell lines was inhibited after exposure to hBD-3, while the proliferation of these cells were not affected by hBD-3. In addition, mutation of defensin-associated genes or cancer-specific loss of human β-defensin expression has been detected in renal cell carcinoma, basal cell carcinoma, and prostate cancer [20,21]. The inhibition of tumor cell growth and apoptosis can be induced by transduction of the defensin gene or treatment with recombinant human β-defensin [22]. Therefore, the researchers proposed that human β-defensin might be a potential suppressor of tumor growth.

Similarly, human β-defensin expression was detected in normal oral cavity mucosa [23], while its expression was decreased in some precancerous lesions, oral cavity cancer tissues and cell lines [24]. Han et al. analyzed the role of human β-defensin in the oral cancer progression and its role as a prognostic and diagnostic parameter and therapeutic target, and showed that human β-defensin inhibited migration and invasion of oral cavity cancer [25]. Our study demonstrated that immune cell infiltration was not significantly different between the N0 and N+ groups. Considering this result, the expression of defensin-related genes from
immune cells around the tongue cancer lesion may be an important parameter rather than the extent of immune cell infiltration, particularly in early stage cancers.

There have been several studies that evaluate saliva samples regarding detecting oral cavity cancer. Cao et al. reported that methylation genomic loci which encodes microRNA biomarkers represents a promising and novel screening tool, and the seven-microRNA panel is able to strongly find oral cavity cancer in saliva [26]. In addition, Gualtero et al. reported the biomarkers in saliva for the detection of oral cancers and their possible utilization for early detection [27]. Considering that defensin can be detected in the saliva of normal and patients with oral cancer [28], we propose that further study can elucidate the role of saliva defensin in tongue cancer diagnosis and prognostic prediction.

Although decreased defensin expression may be associated with cancer metastasis, contrary results have also been reported. Shuyi et al. reported that β-defensin is strongly expressed and significantly associated with lymphatic invasion in oral cancers [29]. In addition, Mburu et al. demonstrated that β-defensin represents an NF-kB-regulated mediator of CCR7 expression and anti-apoptotic signaling, which may be exploited by developing treatments for the patients with head and neck cancer to enhance their survival and reduce metastasis [30]. Our study has limitations that the small sample size in our analysis might decreased validity and credibility of the arrived conclusions. In addition, the external validation was not included in our study which might increase the possibility of bias. Therefore, our results may be not conclusive, and future study is needed to clarify the potential role of defensin expression of tongue cancer metastasis. For this reason, we plan to perform a genetic network study of other to better predict metastasis in a multiple gene expression prediction model.
In conclusion, defensin (DEFB4A, DEFB103B, DEFB4B) may be a novel biomarker for early regional metastasis in T1/2 tongue cancer. However, because the sample size is very small, which in combination to the fact that the differential expression of the genes did not reach significance in the TCGA dataset, further validation should be performed for the clinical usefulness.
REFERENCES


16. Yang D, Biragyn A, Hoover DM, Lubkowski J, Oppenheim JJ. Multiple roles of antimicrobial defensins, cathelicidins, and eosinophil-derived neurotoxin in host


FIGURE LEGENDS

Figure 1. Differentially expressed genes
Among 51 differentially expressed genes, DEFB4A, SPRR2B, DEFB103B, SPRR2G, DEFB4B, and FAM25A were most significantly downregulated genes.

Figure 2. Heatmap

Figure 3. Immune cell infiltration
Immune cell infiltration was analyzed using CIBERSORT program. There was no significant difference in the 22 different types of immune cells and the ratio of M1 to M2 macrophages.

Figure 4. TCGA data analysis
Although statistical significance was not attained, N+ group showed lower level of DEFB103B and DEFB4A.

Figure 5. Immunohistochemical staining
Representative images of defensin staining of each group (upper x100 magnification, lower x200 magnification). The composition of defensin in the cancerous epithelium was significantly higher in the N0 group compared to the N+ group. The intensity of defensin
staining was measured and compared using Image J software. Independent t-test revealed that two groups showed significant difference (p<0.001).
Table 1. Clinicopathologic data of N0 and N+ groups.

<table>
<thead>
<tr>
<th></th>
<th>N0 (n=23)</th>
<th>N+ (n=12)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>58.7±14.8</td>
<td>57.8±22.6</td>
<td>0.877</td>
</tr>
<tr>
<td>Sex (M:F)</td>
<td>15:8</td>
<td>10:2</td>
<td>0.263</td>
</tr>
<tr>
<td>Tumor size (largest, cm)</td>
<td>2.1±1.1</td>
<td>2.3±1.2</td>
<td>0.615</td>
</tr>
<tr>
<td>Tumor thickness (mm)</td>
<td>7.5±2.3</td>
<td>8.3±2.8</td>
<td>0.754</td>
</tr>
<tr>
<td>Nearest resection margin (mm)</td>
<td>3.3±2.0</td>
<td>2.8±2.5</td>
<td>0.545</td>
</tr>
<tr>
<td>Number of metastatic LN</td>
<td>NA</td>
<td>2.0±1.6</td>
<td>-</td>
</tr>
<tr>
<td>Involved LN level*</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN I</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>LN II</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>LN III</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Recurrence</td>
<td>3 (13.0%)</td>
<td>4 (33.3%)</td>
<td>0.163</td>
</tr>
</tbody>
</table>

*not mutually exclusive

LN, lymph node
Table 2. GO enrichment / KEGG pathway analysis

<table>
<thead>
<tr>
<th>Enrichment score</th>
<th>Enrichment</th>
<th>Count</th>
<th>-log(P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.39</td>
<td>Secreted</td>
<td>22</td>
<td>9.886057</td>
</tr>
<tr>
<td></td>
<td>Signal</td>
<td>26</td>
<td>6.744727</td>
</tr>
<tr>
<td></td>
<td>Disulfide bond</td>
<td>22</td>
<td>5.568636</td>
</tr>
<tr>
<td>4.83</td>
<td>Protease inhibitor</td>
<td>7</td>
<td>6.552842</td>
</tr>
<tr>
<td></td>
<td>Serine protease inhibitor</td>
<td>6</td>
<td>5.920819</td>
</tr>
<tr>
<td></td>
<td>Negative regulation of endopeptidase activity</td>
<td>5</td>
<td>3.721246</td>
</tr>
<tr>
<td>4.48</td>
<td>Antibiotic</td>
<td>8</td>
<td>8.853872</td>
</tr>
<tr>
<td></td>
<td>Antimicrobial</td>
<td>8</td>
<td>8.455932</td>
</tr>
<tr>
<td></td>
<td>Beta defensin type</td>
<td>4</td>
<td>6.000000</td>
</tr>
<tr>
<td></td>
<td>Defense response to bacterium</td>
<td>6</td>
<td>4.638272</td>
</tr>
<tr>
<td></td>
<td>Defensin</td>
<td>8</td>
<td>4.207608</td>
</tr>
<tr>
<td>3.59</td>
<td>Keratinocyte differentiation</td>
<td>5</td>
<td>4.508638</td>
</tr>
<tr>
<td></td>
<td>Keratinization</td>
<td>4</td>
<td>4.075721</td>
</tr>
</tbody>
</table>
Figure 1

Gene ID

Log₂ gene expression

DEG (absFC>2)

DEFB4A, SPRR2B, DEFB103B, SPRR2G, DEFB4B, FAM25A
Figure 2

[Image of a heat map with a color key and dendrograms on the sides.]
Figure 3

**Immune cell infiltration**

- N0
- N+

**M1/M2**

- p=0.661
Figure 4

The figure shows a box plot comparing the Log2 gene expression levels of different conditions for DEF103B(N0), DEF103B(N+), DEF4A(N0), and DEF4A(N+). The null hypothesis (ns) indicates no significant difference between the groups.
Figure 5

The diagram shows a comparison of IHC (beta defensin) intensity between the N0 and N+ groups. The bar graph indicates a statistically significant difference with p<0.01. The images of tissue sections demonstrate the intensity levels for each group.