치의학박사 학위논문

# Identification of candidate biomarkers for early diagnosis of oral squamous cell carcinoma

# 구강편평상피세포암종의 조기 진단을 위한 후보 생체지표의 규명

2015년 8월

서울대학교 대학원

치의학과 구강악안면외과학 전공

김 용 덕

**-Abstract-**

# Identification of candidate biomarkers for early diagnosis of oral squamous cell carcinoma

Yong-Deok Kim

Program of Oral and Maxillofacial Surgery,

Department of Dentistry, Seoul National University Graduate School

(Directed by Professor Jong-Ho Lee, D.D.S., M.S.D., Ph.D.)

**Objectives**: Early detection of oral squamous cell carcinoma (OSCC) is critical, because of its rapid growth, aggressive infiltration, frequent lymph node metastasis and poor prognosis. However, clinically valuable tools and biomarkers from systemic study for early diagnosis of OSCC have not been established yet. The purpose of this study was to identify candidate biomarkers for OSCC for early diagnosis with genetic and bioinformatical verification.

**Materials and Methods**: Tumor and normal tissues from patients with OSCC were used for microarray analysis. Differentially expressed genes, identified using permutation, local pooled error (LPE), and t-tests and significance analysis of microarrays (SAM) were investigated with ontological analysis and selected as candidate genetic markers. These genes were validated via quantitative reverse transcription polymerase chain reaction (qRT-PCR).

**Results**: Two groups corresponding with tissue identity were evident, implying that their differentially expressed genes represented the biological differences between tissues. Fifteen genes were identified using Student's t-test ($p<0.05$) and SAM. Based on gene expression, these 15 genes were distinguished in OSCC samples. Genetic analysis of functional networks and ontologies, validated by qRT-PCR of tissue samples, identified four genes, *ADAM15, CDC7, IL12RB2* and *TNFRSF8*, demonstrating favorable concordance with microarray data.

**Conclusion**: *ADAM15, CDC7, IL12RB2* and *TNFRSF8* can be candidate biomarkers for the diagnosis of OSCC.

---

**Keywords** : biomarker, oral squamous cell carcinoma, microarray, quantitative reverse transcription polymerase chain reaction, gene expression profiling.

**Student Number** : 2006-30985

# - Contents -

# I. Introduction

Oral squamous cell carcinoma (OSCC) is the sixth most common malignancy worldwide, accounting for 3% of all cancers (1,2), and it remains one of the most intractable malignancies due to its invasive growth pattern, frequent cervical lymph-node metastasis and high recurrence rate (3). Approximately two-thirds of patients with OSCC exhibit an advanced stage (Stage III or IV) at diagnosis because lesions are often asymptomatic and clinicians cannot predict the biological relevance of lesions on the basis of their physical features alone (4,5). As OSCC is associated with severe morbidities and a low 5-year survival rate (approximately 50%), especially for advanced-stage disease (41% for Stage III and 9% for Stage IV), early detection is critical (6,7). For that purpose, screening methods, such as light-based screening or brush cytology, have been introduced, but their sensitivity and specificity are insufficient compared with those of scalpel biopsy. Hence, histopathological examination is still a cornerstone of diagnosis; however, excluding the possibility of misdiagnosis completely is difficult due to the inherent vagueness of the histopathological characteristics of OSCC and variations in the

pathologists' experience.

Recently, mRNA biomarkers for OSCC in serum or tissue have become new diagnostic and therapeutic targets because suitable biomarkers can improve the power, availability and cost-effectiveness of high-throughput screening for genetic alterations. Microarray analysis, validated with the quantitative reverse transcriptase polymerase chain reaction (qRT-PCR), has been used to identify the genes underlying OSCC pathogenesis, which include *IL-8*, vascular endothelial growth factor (*VEGF*) (8), laminin γ2 (*LAMC2*) / collagen type IV α1 (*COL4A1*) and collagen type I α1 (*COL1A1*) / peptidyl arginine deiminase, type I (*PADI1*) (9), and the homeobox (HOX)-genes model *HOXA5* / *HOXD10*/ *HOXD11* (10). Some investigators have employed a multiple-gene model of 25 biomarkers with 86%-89% accuracy (11). However, no studies have established clinically-valuable biomarkers, which may be due to a lack of studies with tumor-normal (TN) paired matching of patients (10) or to the limited predictive power of microarray-based models to correlate the clinical endpoint with gene expression (12). This study is aimed to identify the candidate biomarkers of OSCC using TN- paired matching samples, verified by bioinformatical methods with reduced statistical error.

# II. Materials & Methods

## Patients

Tumor and normal tissues from five OSCC patients were used for the microarray analysis. Subsequently, tissues from 17 OSCC patients (TN-paired) were used for the qRT-PCR analysis. Fresh tissue samples were collected after obtaining written informed consent from 33 consecutive patients undergoing therapeutic surgical resection for OSCC between 1 June 2011 and 28 June 2012. The patients' characteristics are outlined in table 1.

After excision, tissues were preserved immediately in RNAlater® solution (Invitrogen, Waltham, MA, USA) and were transferred to the laboratory on ice. This trial was approved by the Institutional Review Board at Seoul National University Dental Hospital and was conducted in full accordance with the Declaration of Helsinki.

## RNA extraction and cDNA synthesis

RNAlater® was pipetted off the pellet, and the pellet was then washed with ice-cold

phosphate-buffered saline, which was removed after centrifugation. Total RNA was extracted from the tissue samples with the RNeasy® Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. To quantify and analyze the integrity of RNA, 2100 Bioanalyzer® (Agilent Technologies, Santa Clara, CA, USA) was used. 10 μg of total RNA was reversely transcribed in the presence of an oligo (dT) T7 primer (iScript™ Select cDNA Synthesis Kit; Bio-Rad Laboratories, Hercules, CA, USA). cDNA was used for in-vitro transcription amplification in the presence of biotinylated nucleotides.

## Microarray analysis

Gene expression was analyzed using the Affymetrix preparation protocol before hybridization to a Gene-Chip® 1.1 human genome microarray (Affymetrix, Santa Clara, CA, USA). Quality control of the arrays was achieved through analysis of the 5':3' ratios (range: 0.40-0.79), the percentage present (range: 37-47%), the average pair-wise correlation, and the principal component. Affymetrix ® Microarray Suite version 5.0 (Affymetrix, Santa Clara, CA, USA) was used for image processing and data acquisition.

Gene expression levels and individual exon signal estimates contained in *. CEL files within the GeneChip® 1.1 platform (Affymetrix, Santa Clara, CA, USA) were derived by the robust multi-array average (RMA) algorithm, as implemented by the Expression Console v1.1.1 (Affymetrix, Santa Clara, CA, USA). The quality of the data was assessed by analyses of the mean overall expression and its standard deviation, the detectable probe ratios, the control probe's (housekeeping genes) expression data, and the correlation between samples from the same tissue.

## qRT-PCR

Target and housekeeping gene sequences were retrieved from the GenBank® database (National Center for Biotechnology Information (NCBI), Bethesda, MD, USA), and applied primers were manually designed using the Primer-BLAST® tool (NCBI, Bethesda, MD, USA). Primer sequences are listed in Table 2. RNA mixtures were subjected to first-strand cDNA synthesis using forward primers. The endogenous housekeeping genes, hypoxanthine guanine phosphoribosyl transferase (*HPRT*) and β–actin (*ACTB*), were used for data normalization, and relative quantification was performed with a relative standard

curve analysis with a CFX Connect™ Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) and SYBR® Green I (Bio-Rad Laboratories, Hercules, CA, USA) detection. cDNA (10ng) was dissolved in iQ™SYBR® Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA). The PCR comprised an initialization cycle (95°C for 3 minutes); 39 cycles of 95°C for 10 seconds, 56°C for 10 seconds, 72°C for 20 seconds, and plate reading at 80°C; and a final melting step at temperatures ranging from 56 to 95°C at a heating rate of 0.5°C/10 seconds, which was performed to create melt curves. The qRT-PCR was performed in quadruplicate for each cDNA samples, and negative controls with no template were included for each primer pair. Threshold cycle (Ct) values and target gene expression levels were calculated using CFX Manager™ Software ((Bio-Rad Laboratories, Hercules, CA, USA). The fold change of target gene expression in each treated sample, relative to the control sample, was derived from the negative Ct value: Ct (GAPDH)−Ct (target gene).

## Data analysis and prediction models

Differences in the clinical characteristics between the Microarray Study and the qRT-

PCR Validation groups were compared using the Chi-square test. If this method was not suitable, a Fisher's exact test or a linear-by-linear association test was used. Differentially expressed genes (DEGs), identified with four statistical tests (permutation test, local pooled error (LPE) (13), Student's t-test, and significance analysis of microarrays (SAM) (14), were selected as candidates for genetic biomarkers (15), and their normalities were tested using the Shapiro-Wilk test (16). After the normality had been tested, genes that were normally distributed were assessed using the Student's t-test, the SAM and the LPE (parametric methods). The genes that were not distributed normally were assessed using a non-parametric permutation test. For the multiple-comparison problem, p-values were adjusted using the false discovery rate (FDR) correction. To find candidates, FDR values as close to 0% as possible were used , and those for the SAM and the permutation test were set to 0%. On the other hand, the FDRs for the LPE and the Student's t-test were set to 0.001% and 0.02%, respectively, because no significant candidate genes were found when their FDRs were set to 0%. A discriminant analysis was used to find correlations between the candidate gene expressions and clinical parameters such as tissue heterogeneity, site specificity, and stage differences. From receiver-operating

characteristic (ROC) curves, the area under the curve (AUC), the sensitivity, the specificity, and the predictive value of each candidate gene were calculated using the Youden index. All statistical analyses, except the analyses of the ROC curves, which were done using Med-Calc® Statistical Software version 12.7.8 (MedCalc® Software bvba, Oostende, Belgium), were conducted using R Statistical Software version 2.15.0 (from http://www.r-project.org/) and IBM SPSS® version 20 (IBM Co., Armonk, NY, USA).

# IV. Results

## Patient characteristics

Thirty-three patients with OSCC, 17 men (51.5%) and 16 women (48.5%), were included in this study (Table 1). The DEGs in five patients were analyzed with microarray analysis to compare primary TN-paired tissue samples. Tumor (n=20) and normal tissues (n=25) from another 28 patients were used for the qRT-PCR validation. Of those 28 patients, 17 were assayed for expression via the qRT-PCR of primary tumor samples and matched normal mucosa. No statistically significant differences in clinical characteristics between the microarray study and the qRT-PCR validation groups were observed ($p > 0.05$).

## Statistical analysis of candidate genes identified with a microarray analysis

All data were found to be followed a normal distribution in Shapiro-Wilk test. Therefore, parametric methods (the Student's t-test, the SAM and the LPE) were used to identify DEGs. The LPE test was proposed to overcome the limitation caused by the

different error variances arising from the diverse biological conditions via local error estimation within quantiles and non-parametric smoothing (13). The SAM scores gene expression relative to the deviation of repeated measurements in order to identify genes with high scores above an adjustable threshold, and for those identified genes, the FDRs and the percentage of identification by chance were estimated (Figure 1) (14). Of the 33,297 genes on the microarray, the numbers of genes with significance, which were assessed with the Student's t-test, the SAM and the LPE (parametric methods), were 779, 349 and 200, respectively. None of these genes showed statistical significance in all 3 methods, but fifteen genes passed the Student's t-test ($p$ <0.05) and the SAM analysis. To test the ability of this signature gene set to classify the OSCC and the Normal groups, an average linkage hierarchical clustering analysis was performed. Based on gene expression, this 15-gene set had classification power for OSCC (Figure 2).

In this set, upregulated genes were family with sequence similarity 213, member B (C1orf93, *FAM213B*), Rho guanine nucleotide exchange factor 16 (*ARHGEF16*), zinc finger and BTB domain containing 48 (*ZBTB48*), DNA cross-link repair 1B (*DCLRE1B*), late cornified envelope 1D (*LCE1D*), kelch domain containing 7A (*KLHDC7A*), Tumor necrosis factor receptor superfamily, member 8 (*TNFRSF8*), Interleukin 12 receptor, beta 1 (*IL12RB2*), Cell division cycle 7 (*CDC7*) and A disintegrin and metallaproteinase metallopeptidase domain 15 (*ADAM15*), while UTP11-like, U3 small nucleolar ribonucleoprotein (*UTP11L*), polo-like kinase 3

(*PLK3*), WD repeat domain 63 (*WDR63*), potassium voltage-gated channel, Shaw-related subfamily, member 4 (*KCNC4*) and ATPase, aminophospholipid transporter, class I, type 8B, member 2 (*ATP8B2*) were downregulated (Table 3).

## Analysis of the literature related to candidate genes

An ontological analysis revealed the function and the disease association of each gene, identifying genes involved in inflammation or immunity (*FAM123B, ZBTB48, TNFRSF8 and IL12RB2*), intercellular signaling and movement (*ARGHEF16, KCNC4 and ADAM15*), transcription (*UTP11L and ATP8B2*), cell division (*PLK3, CDC7 and DCLRE1B*) and keratinization (*LCE1D*) (17-27). After a literature review, well-known genes associated with carcinogenesis, such as *PLK 3* and *KCNC 4*, and unidentified genes, including *WDR63* and *KLHDC7A* were excluded. Among the remaining genes, upregulated ones were chosen as candidates. Of them, *ARHGEF16* and *DCLRE1B* were excluded due to their unknown correlation with oral cancer. Finally, *IL12RB2, TNFRSF8, ADAM15* and *CDC7* remained. According to previous research, these 4 genes are known to be related to inflammation, carcinogenesis or metastasis (17,19,21,22).

## qRT-PCR validation of gene expression

Expressions of the four candidate genes, *ADAM15, CDC7, IL12RB2* and *TNFRSF8*, were assayed with the qRT-PCR in 28 OSCC samples to correlate microarray data with the qRT-PCR data and to explore potential biomarker expression (Figure 3). Relative expression levels of the four genes were obtained by normalization to housekeeping gene expression levels. The qRT-PCR expression values for the four candidate genes, whose expression levels were higher in tumor tissues compared with normal tissues, were used for correlation analyses using the 2-ΔΔCt method with the Student's t-test. In the tumor group, statistically-significant high mRNA expression levels of *ADAM15* and *CDC7* were consistently detected with the qRT-PCR ($p<0.001$ and $p<0.001$, respectively), thus showing a good correlation with the microarray data. A higher expression in the tumor relative to normal tissue was particularly evident for *CDC7* in its increased 2-ΔΔCt value.

The mRNA expression of the selected genes in 17 TN-paired tissue samples was also measured with the qRT-PCR. The normalized expression was calculated with the 2-ΔΔCt method, as previously described, to compare normal and tumor tissues. *ADAM15* expression was significantly increased in tumor tissues ($p<0.001$) by an average of 2.7-

fold compared with normal tissues (Figure 4). Consistent with these findings, upregulation of *CDC7* was evident in tumor tissues ($p < 0.05$). Furthermore, the candidate genes were overexpressed regardless of variability in the degree of expression between samples in the TN-paired group.

## Evaluation of the diagnostic validity of candidate genes

The relationships between the clinical parameters and the candidate genes are summarized in Table 4. Classification of the primary sites with all four candidate genes as signatures resulted in moderate accuracy (64.7%) than using the *ADAM15* (35.3%), *CDC7* (47.1%), *IL12RB2* (23.5%), and *TNFRSF8* (41.2%) by themselves. Similar tendency was also observed regarding tissue heterogeneity and staging. These results imply that a combination of multiple candidate genes has a discriminatory power that is superior to that achieved with each candidate gene alone. ROC plots are presented in Figure 5, and their diagnostic indices were shown in Table 5. The AUCs of *ADAM15, CDC7, IL12RB2* and *TNFRSF8* were 0.699, 0.645, 0.561 and 0.602, respectively.

*ADAM15* alone, however, showed significance ($p < 0.05$). Similarly, the odds ratio of *ADAM15* (7.271; $p = 0.029$) and *CDC7* (26.794; $p = 0.046$) were significant.

# V. Discussion

The five-year survival rate of patients with OSCC has remained about 50% for several decades. This is one of the worst prognoses of all cancers and is directly related to the advanced stage at diagnosis (28). For the early diagnosis of OSCC, biomarkers should be minimally invasive as possible with simple procedure and rapid result. Biomarkers of OSCC can be DNA, mRNA, proteins, metabolites, or process such as apoptosis, angiogenesis or proliferation in tissue, saliva and serum. Biomarkers from body fluid are non invasive and easy to access, but their accuracy and sensitivity are not reliable and preicise localization of lesion is difficult. As OSCC is multi-factorial with heterogenic pathogenesis, combination of mRNA or protein biomarkers with different sources is necessary for the discrimination between tumor and controls (29,30). Thus identification of candidate mRNA biomarkers is the fundamental for the establishment of systemic OSCC biomarkers. Moreover, local recurrence and distant metastasis may develop despite an apparently complete excision and histopathologically tumor-free margins, supporting the theory of field cancerization, defined as malignant or premalignant changes in an entire field of apparently normal tissue surrounding a tumor in response to

carcinogens (31). mRNA biomarkers could allow a more sensitive and accurate surgical margin to be determined. However, genome research has been hindered by individual and ethnic differences (32). Identification of OSCC-specific biomarkers in the populations is therefore necessary.

Although the sample size of the microarray study group (n = 5) seems too small, accurate result can be ensured in in three ways: (i) several statistical analyses (the permutation test, Student's t-test, the LPE and the SAM), which can identify significant DEGs with a small number of microarrays, were applied (13), (ii) genes carefully selected with these statistical tools were validated via the qRT-PCR with sufficient samples, and (iii) as shown in Table 1, patients' clinicopathological characteristics were evenly distributed in both the microarray study and the qRT-PCR Validation groups. In spite of a general consensus that smoking history is a major contributor to carcinogenesis (33), gene expression between smoking and non-smoking subjects did not differ in this study. Further, Méndez *et al*. observed no significant differences in gene expression between patients with different histopathologic stages when using oligonucleotide arrays (34). Thus, the effect of smoking history and histopathological type on the results can be

expected to be minimal.

To date, many studies have attempted to identify candidate biomarkers for OSCC (3,8-10). However, they were not generally accepted for the following reasons: First, candidate biomarkers yielded by the microarray analysis were usually identified based on fold changes (9), but these can often be misleading, as different error variances exist under different biological conditions and in different microarray expression intensity ranges (13). Second, most previous studies mainly focused on well-known genes, such as matrix metalloproteinases (*MMPs*), *p53* and *VEGF* (8). Such an approach could result in a bias against unknown genes, so selection of candidates with a microarray anlysis which is a relatively unbiased high-throughput method and reducing the statistical error with four kinds of biostatistics verification can overcome such prejudices.

From these results, 15 genes were initially screened (Table 3). After exclusion of genes with unknown functions as previously mentioned, eight genes (*ARHGEF16, DCLRE1B, TNFRSF8, PLK3, IL12RB2, CDC7, KCNC4* and *ADAM15*) are reportedly associated with cancer. Particularly, *PLK3* and *KCNC4* are well-known DEGs in head and neck cancer. *PLK3* is an important mediator of the cellular responses to genotoxic

stresses and known to regulate the activity of *p53* (35). *PLK3* down-regulation in OSCC tissue is consistent with results of studies of both head and neck and esophageal SCC (23,36), suggesting that *PLK3* downregulation functions in SCC tumorigenesis. *KCNC4*, encoding components of voltage-gated potassium channels, is reportedly overexpressed in head and neck cancer (32) ; however, it was downregulated in OSCC tissue in this study. Tumor markers secreted by tumor cells or released subsequently to tumor-cell fragmentation are valuable, because they are quantitatively more highly expressed than downregulated genes, and thereby increase the sensitivity of cancer diagnosis (37), thus *PLK3* and *KCNC4* were excluded. As illustrated in Figure 2, upregulated genes whose functions are related to cancer comprise *ARHGEF16, DCLRE1B, TNFRSF8, IL12RB2, CDC7* and *ADAM15*. However, few studies report an association between *ARHGEF16* and *DCLRE1B* in head and neck malignancies. Therefore, the remaining four genes (*ADAM15, CDC7, IL12RB2* and *TNFRSF8*) were selected for further evaluation using qRT-PCR.

*ADAM15* is the only member of a disintegrin and metalloproteinase (ADAM) family with an RGD (Arg-Gly-Asp) binding motif, enabling it to bind to integrins (38) and is

involved in T-cell-mediated immune responses and pathologic neovascularization, implicated in several processes associated with tumor progression and inflammation (39). It is upregulated in breast, stomach, lung, pancreas, and prostate cancer and in metastatic progression (40). Conversely, its tumor suppression in melanomas and colon carcinomas was also proposed (38,41). This inconsistency might result from the complexity of the *ADAM15*-related signal transduction pathways. In this study, *ADAM15* was significantly upregulated in patients with OSCC. *CDC7*, a serine threonine kinase, is a core component of the initiation machinery of DNA replication. Its activity is essential for S phase progression and inhibition of *CDC7* in cancer cells induces *p53*-independent apoptosis (42). Clinically, increased *CDC7* expression was found to be linked to tumor anaplasia, aneuploidy, advanced clinical stage and lower relapse-free survival, implicating *CDC7* dysregulation in the development of aggressive disease in ovarian cancer, breast cancer, large B-cell lymphoma and malignant melanoma (42-45). Studies of human papilloma virus-related head and neck SCC demonstrated high expression of *CDC7*, similarly to other cell cycle regulators, such as cyclin-dependent kinase inhibitor 2A and cyclin-dependent kinase inhibitor 2C (45). These results agreed with the data, which showed

overexpression of *CDC7* in patients with OSCC. *IL12RB2* encodes a type-1 transmembrane protein, a subunit of the Interleukin 12 (*IL-12*) receptor complex: the cytokine *IL-12* bridges innate and adaptive immunity and regulates T-cell, natural killer cell and B-cell function. *IL12RB2* behaves as a tumor suppressor in human chronic B-cell malignancies, and its silencing is an early event in B-cell malignant transformation (32,46). Prolonged disease-free survival has also been reported to have been found in colorectal cancer patients with higher expressions of *IL12RB2* (47). However, the correlation between *IL12RB2* expression and OSCC development is unclear, a case of oesophageal squamous cell carcinoma in young adult with *IL12RB1* deficiency was reported (48). Moreover, considering that interleukin 6 (IL-6) and IL-12 functionally antagonize each other and that IL-6-induced inflammation promotes carcinogenesis in the oral cavity (49,50), *IL12RB2* downregulation might be involved in OSCC. However, despite overexpression of *IL12RB2* in patients with OSCC in the microarray analysis group, no significant differences of *IL12RB2* expression were found in this qRT-PCR data. *TNFRSF8* is known to function in apoptosis by activating the caspase pathway (51) and protecting the body against autoimmunity. *TNFRSF8* expression is upregulated in

various hematological malignancies, including Reed–Sternberg cells in Hodgkin's disease, anaplastic large-cell lymphoma and subsets of non-Hodgkin's lymphoma (52). It is a conventional biomarker of hematologic malignancies such as anaplastic large-cell lymphoma (53), but its relationship with solid tumors is unclear. Only one study reported that its overexpression might be associated with improved prognosis for patients with a skull base chordoma (54). Though *TNFRSF8* in patients with OSCC was upregulated in the microarray analysis, qRT-PCR validation showed no significant differences.

Among the four candidates, *IL12RB2* and *TNFRSF8* are known to be related to inflammation. Given that the oral cavity is vulnerable to frequent inflammatory stress and that this environment can affect cancer cells, these results support that inflammation might be correlated with OSCC development, as reported in some recent studies (8,50). Furthermore, the correlationship between OSCC and oral lichen planus (OLP) and graft-versus-host disease (GVHD) can be deduced. OLP is a chronic inflammatory disease as a result of the T-cell-mediated autoimmune response and known as a premalignant lesion of OSCC (55). *IL12RB2* and *TNFRSF8* expression in OLP tissues were compared with non-OLP (mucositis) samples and *TNFRSF8* was upregulated in OLP tissues as

similar to OSCC samples in the microarray results (56). These two genes are also related to GVHD, which has similar clinical and histopathological features to OLP (57). Chronic GVHD (cGVHD) has long been suspected to be a potential risk factor for secondary malignancies and several authors have detected OSCC in patients with extensive cGVHD (58). Taken together, OLP and GVHD seem to be associated with OSCC and these findings may contribute to identify therapeutic targets for OSCC, but further researches should be warranted.

The correlation between the candidate genes and the clinical parameters (primary site, tissue heterogeneity and stage) were presented in Table 4. Although all *p* values were higher than 0.05, the eigenvalues and the accuracies of the signatures of the four genes for each parameter were generally lower than those of individual genes according to the Wilk's lambda distribution analysis. This result indicates that the classifying power of the four genes combined was superior to that of each gene alone.

As presented in table 5, the sensitivity, specificity and predictive values were calculated at the cut-off points according to the Youden index. At the cut-off points, the value of [sensitivity + specificity - 1] is a maximum (59). The Youden index is widely used to

determine the diagnostic threshold, and it is regarded as the most appropriate method of combining biomarkers (60). In terms of the AUC, *ADAM15* showed higher clinical efficacy ($p < 0.05$) compared to the other candidate genes. The odds ratios of *ADAM15* and *CDC7* also showed significant correlations to OSCC ($p < 0.05$).

For further evalutions, the genetic product of candidates would be investigated with immunohistochemistry or western blot methods. Next generation sequencing could be a new method for the validation of candidate genes in genomic level also. But the dimension of the samples in this study were around 1cm × 1cm, so it was difficult for further verification process after qRT-PCR analysis in same tissues. Large samples from different status and parameters, especially of the cancers developed from premalignant disease could contribute a variety of bioinformations for delicate diagnostic biomarkers, but serial sample collections of precancerous lesion from same patient are not easy to be obtained. Thus continuous investigation with diverse samples should be required to validate the clinical feasibility of candidates for the early diagnosis (61).

To conclude, *ADAM15, CDC7, IL12RB2 and TNFRSF8* might be valuable candidate biomarkers for the early diagnosis of OSCC and further study should be necessary for the

establishment of systemic biomarkers for diagnosis and treatment of OSCC.

# VI. References


1. Parkin DM. Global cancer statistics in the year 2000. *Lancet Oncol*. 2001;2:533-43.
2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer statistics, 2008. *CA Cancer J Clin.* 2008;58:71-96.
3. da Silva SD, Ferlito A, Takes RP, Brakenhoff RH, Valentin MD, Woolgar JA, et al. Advances and applications of oral cancer basic research. *Oral Oncol*. 2011;47:783-91.
4. Rethman MP, Carpenter W, Cohen EE, Epstein J, Evans CA, Flaitz CM, et al. Evidence-based clinical recommendations regarding screening for oral squamous cell carcinomas. *J Am Dent Assoc.* 2010;141(5):509-20.
5. McLeod NM, Saeed NR, Ali EA. Oral cancer: Delays in referral and diagnosis persist. *Br Dent J.* 2005;198:681-84.
6. Neville BW. Oral and Maxillofacial Pathology, Saunders/Elsevier**:** 2002; 356-67.
7. Mao L, Hong WK, Papadimitrakopoulou VA. Focus on head and neck cancer. *Cancer Cell.* 2004; 5(4): 311-16.
8. Rao SK, Pavicevic Z, Du Z, Kim JG, Fan M, Jiao Y, et al. Pro-inflammatory genes as biomarkers and therapeutic targets in oral squamous cell carcinoma. *J Biol Chem.* 2010; 285:32512-21.
9. Chen C, Méndez E, Houck J, Fan W, Lohavanichbutr P, Doody D, et al. Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2008;17:2152-62.
10. Rodini CO, Xavier FC, Paiva KB, De Souza MF, Moyses RA, Michaluarte P, *et al*. Homeobox gene expression profile indicates *HOXA5* as a candidate prognostic marker in oral squamous cell carcinoma. *Int J Oncol* . 2012;40(4): 1180-8.
11. Ziober AF, Patel KR, Alawi F, Gimotty P, Weber RS, Feldman MM, et al. Identification of a gene signature for rapid screening of oral squamous cell carcinoma. *Clin Cancer Res*. 2006;12:5960-71.
12. Zhao J, Rycaj K, Geng S, Li M, Plummer JB, Yin B, et al. Expression of human endogenous retrovirus type K envelope protein is a novel candidate prognostic marker for human breast vancer. *Genes Cancer*. 2011;2:914-22.
13. Jain N, Thatte J, Braciale T, Ley K, O'Connell M, Lee JK. Local- pooled-error test for

identifying differentially expressed genes with a small number of replicated microarrays. *Bioinformatics*. 2003;19:1945-51.

14. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci*. 2001;98:5116-21.
15. Lockstone HE. Exon array data analysis using Affymetrix power tools and R statistical software. *Brief Bioinform*. 2011;12:634-44.
16. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52:591-611.
17. Dürkop H, Latza U, Hummel M, Eitelbach F, Seed B, Stein H. Molecular cloning and expression of a new member of the nerve growth factor receptor family that is characteristic for Hodgkin's disease. *Cell*. 1992;68:421-7.
18. Li B, Ouyang B, Pan H, Reissmann PT, Slamon DJ, Arceci R, et al. *Prk*, a cytokine-inducible human protein serine/threonine kinase whose expression appears to be down-regulated in lung carcinomas. *J Biol Chem*. 1996;271:19402-8.
19. Bassuny WM, Ihara K, Kimura J, Ichikawa S, Kuromaru R, Miyako K, et al. Association study between interleukin-12 receptor beta1/beta2 genes and type 1 diabetes or asthma in the Japanese population. *Immunogenetics.* 2003;55:189-92.
20. Harris MJ, Arias IM. *FIC1,* a P-type ATPase linked to cholestatic liver disease, has homologues (*ATP8B2* and *ATP8B3*) expressed throughout the body. *Biochim Biophys Acta*. 2003;1633:127-31.
21. Tenca P, Brotherton D, Montagnoli A, Rainoldi S, Albanese C, Santocanale C. *Cdc7* is an active kinase in human cancer cells undergoing replication stress. *J Biol Chem*. 2007;282:208-15.
22. Zhong JL, Poghosyan Z, Pennington CJ, Scott X, Handsley MM, Warn A, et al. Distinct functions of natural *ADAM-15* cytoplasmic domain variants in human mammary carcinoma. *Mol Cancer Res*. 2008;6:383-94.
23. Menéndez ST, Rodrigo JP, Allonca E, García-Carracedo D, Alvarez- Alija G, Casado-Zapico S, et al. Expression and clinical significance of the Kv3.4 potassium channel subunit in the development and progression of head and neck squamous cell carcinomas. *J Pathol*. 2010;221:402-10.
24. Kuwano Y, Kamio Y, Kawai T, Katsuura S, Inada N, Takaki A, et al. Autism-associated gene expression in peripheral leucocytes commonly observed between subjects with

autism and healthy women having autistic children. *PLOS one*. 2011;6:e24723.

25. Marenholz I, Rivera VA, Esparza-Gordillo J, Bauerfeind A, Lee-Kirsch MA, Ciechanowicz A, et al. Association screening in the Epidermal Differentiation Complex (EDC) identifies an SPRR3 repeat number variant as a risk factor for eczema. *J Invest Dermatol.*2011;131:1644-9.
26. Oliver AW, He X, Borthwick K, Donne AJ, Hampson L, Hampson IN. The HPV16 E6 binding protein Tip-1 interacts with *ARHGEF16*, which activates *Cdc42*. *Br J Cancer.* 2011;104:324-31.
27. Lenain C, Bauwens S, Amiard S, Brunori M, Giraud-Panis MJ, Gilson E. The Apollo 5' exonuclease functions together with *TRF2* to protect telomeres from DNA repair. *Curr Biol*. 2006;16:1303-10.
28. Pettus JR, Johnson JJ, Shi Z, Davis JW, Koblinski J, Ghosh S, et al. Multiple kallikrein (KLK 5, 7, 8, and 10) expression in squamous cell carcinoma of the oral cavity. *Histol Histopathol*. 2009;24:197- 207.
29. Yakob M, Fuentes L, Wang MB, Abemayor E, Wong DT. Salivary biomarkers for detection of oral squamous cell carcinoma – current state and recent advances. *Curr Oral Health Rep*. 2014; 1;1(2):133-41.
30. Ni YH, Ding L, Hu QG, Hua ZC. Potential biomarkers for oral squamous cell carcinoma: proteomics discovery and clinical validation. *Proteomics Clin Appl.* 2015;9(1-2):86-97
31. Supic G, Kozomara R, Jovic N, Zeljic K, Magic Z. Prognostic significance of tumor-related genes hypermethylation detected in cancer-free surgical margins of oral squamous cell carcinomas. *Oral Oncol*. 2011;47:702-8.
32. Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, et al. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res.* 2009;19:1622-9.
33. Cheong SC, Chandramouli GV, Saleh A, Zain RB, Lau SH, Sivakumaren S, et al. Gene expression in human oral squamous cell carcinoma is influenced by risk factor exposure. *Oral Oncol*. 2009;45:712-9.
34. Méndez E, Cheng C, Farwell DG, Ricks S, Agoff SN, Futran ND, et al. Transcriptional expression profiles of oral squamous cell carcinomas. *Cancer*. 2002;95:1482-94.
35. Yang Y, Bai J, Shen R, Brown SA, Komissarova E, Huang Y, et al. Polo-like kinase 3 functions as a tumor suppressor and is a negative regulator of hypoxia-inducible factor-

1 alpha under hypoxic conditions *Cancer Res*. 2008;1;68(11):4077-85.

36. Takai N, Hamanaka R, Yoshimatsu J, Miyakwa I. Polo-like kinases (Plks) and cancer. *Oncogene*. 2005; 24(2): 287-291.
37. Yang Y, Pospisil P, Iyer LK, Adelstein SJ, Kassis AI. Integrative genomic data mining for discovery of potential blood-borne biomarkers for early diagnosis of cancer. *PLOS one*. 2008;3:e3661.
38. Ungerer C, Doberstein K, Bürger C, Hardt K, Boehncke WH, Böhm B, et al. *ADAM15* expression is downregulated in melanoma metastasis compared to primary melanoma. *Biochem Biophys Res Commun*. 2010;401:363-9.
39. Lucas N, Najy AJ, Day ML. The therapeutic potential of *ADAM15*. *Current pharmaceutical design*. 2009;15:2311-8.
40. Duffy MJ, McKiernan E, O'Donovan N, McGowan PM. Role of *ADAM*s in cancer formation and progression. *Clin Cancer Res*. 2009;15:1140-4.
41. Toquet C, Colson A, Jarry A, Bezieau S, Volteau C, Boisseau P, et al. *ADAM15* to alpha5beta1 integrin switch in colon carcinoma cells: a late event in cancer progression associated with tumor dedifferentiation and poor prognosis. *Int J Cancer*. 2012;130:278-87.
42. Montagnoli A, Moll J, Colotta F. Targeting cell division cycle 7 kinase: a new approach for cancer therapy. *Clin Cancer Res*. 2010;16:4503-8.
43. Kulkarni AA, Kingsbury SR, Tudzarova S, Hong HK, Loddo M, Rashid M, et al. *Cdc7* kinase is a predictor of survival and a candidate therapeutic target in epithelial ovarian carcinoma. *Clin Cancer Res*. 2009;15:2417-25.
44. Rodriguez-Acebes S, Proctor I, Loddo M, Wollenschlaeger A, Rashid M, Falzon M, et al. Targeting DNA replication before it starts: *Cdc7* as a therapeutic target in p53-mutant breast cancers. *Am J Patho*. 2010;177:2034-45.
45. Slebos RJ, Yi Y, Ely K, Carter J, Evjen A, Zhang X, et al. Gene expression differences associated with human papillomavirus status in head and neck squamous cell carcinoma. *Clin Cancer Res*. 2006;12:701-9.
46. Pistoia V, Cocco C, Airoldi I. Interleukin-12 receptor beta2: from cytokine receptor to gatekeeper gene in human B-cell malignancies. *J Clin Oncol*. 2009;27:4809-16.
47. Tosolini M, Kirilovsky A, Mlecnik B, Fredriksen T, Mauger S, Bindea G, et al. Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, th2, treg,

th17) in patients with colorectal cancer. *Cancer Res*. 2011;71:1263-71.

48. Cardenes M, Angel Moreno A, Fieschi C, Soloquren I, Colino E, Molines A, et al. Oesophageal squamous cell carcinoma in a young adult with IL-12R beta 1 deficiency. *J Med Genet.* 2010;47(9):635-7
49. Airoldi I, Di Carlo E, Cocco C, Sorrentino C, Fais F, Cilli M, et al. Lack of *Il12rb2* signaling predisposes to spontaneous autoimmunity and malignancy. *Blood*. 2005;106:3846-53.
50. Gasche JA, Hoffmann J, Boland CR, Goel A. Interleukin-6 promotes tumorigenesis by altering DNA methylation in oral cancer cells. *Int J Cancer*. 2011;129:1053-63.
51. Kang MR, Kang JS, Yang JW, Kim BG, Kim JA, Jo YN, et al. Gene expression profiling of *KBH-A42*, a candidate histone deacetylase inhibitor, in human leukemia and bladder cancer cell lines. *Oncol Lett*. 2012;3:113-8.
52. Oflazoglu E, Grewal IS, Gerber H. Targeting *CD30/CD30L* in oncology and autoimmune and inflammatory diseases. *Adv Exp Med Biol*. 2009;647: 174-85.
53. Agnelli L, Mereu E, Pellegrino E, Limongi T, Kwee I, Bergaggio E, et al. (European T-Cell Lymphoma Study Group) Identification of a three-gene model as a powerful diagnostic tool for the recognition of ALK negative ALCL. *Blood*. 2012;120:1274-81.
54. Longoni M, Orzan F, Stroppi M, Boari N, Mortini P, Riva P. Evaluation of *1p36* markers and clinical outcome in a skull base chordoma study. *Neuro Oncol*. 2008;10:52-60.
55. Pereira JS, Monteiro BV, Nonaka CF, Silveira EJ, Miguel MC. FoxP3(+) T regulatory cells in oral lichen planus and its correlation with the distinct clinical appearance of the lesions. *Int J Exp Pathol* 2012; 93: 287-94.
56. Jeon SH. Potential of interleukin 12, beta 2 subunit (*IL12RB2*) and tumor necrosis factor superfamily, member 8 (*TNFRSF8*) gene as diagnostic biomarkers of oral lichen planus. Doctor's thesis of Seoul National University. 2014
57. Al-Hashimi I, Schifter M, Lockhart PB, Wray D, Brennan M, Migliorati CA, et al. Oral lichen planus and oral lichenoid lesions: diagnostic and therapeutic considerations. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2007; 103: Suppl. (S25): e1-12.
58. Montebugnoli L, Gissi DB, Marchetti C, Foschini MP. Multiple squamous cell carcinomas of the oral cavity in a young patient with graft-versus-host disease following allogenic bone marrow transplantation. *Int J Oral Maxillofac Surg* 2011; 40:

556-8.

59. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32-5.
60. Yin J, Tian L. Optimal linear combinations of multiple diagnostic biomarkers based on Youden index. *Stat Med*. 2014;33:1426-40.
61. Mahfouz ME, Rodrigo JP, Takes RP, Elsheikh MN, Rinaldo A, Brakenhoff RH, et al. : Current potential and limitations of molecular diagnostic methods in head and neck cancer. *Eur Arch Otorhinolaryngol* 2010;267:851-60.

# Table

**Table 1. Clinicopathologic characteristics of patients and the qualitative reverse transcription polymerase chain reaction validation group.**

| Parameter | Study (Microarray) | Validation (qRT-PCR) | | |
|---|---|---|---|---|
| | TN paired (n=5) | Normal (n=20) | Tumor (n=25) | TN paired (n=17) |
| Age (yrs) | | | | |
| <50 | 0 (0) | 2 (10) | 4 (16) | 2 (11.8) |
| ≥50 | 5 (100) | 18 (90) | 21 (84) | 15 (88.2) |
| Gender | | | | |
| Male | 4 (80) | 10 (50) | 12 (48) | 9 (52.9) |
| Female | 1 (20) | 10 (50) | 13 (52) | 8 (47.1) |
| Therapy prior to surgery | | | | |
| None | 3 (60) | 17 (85) | 21 (84) | 14 (82.4) |
| Yes | 2 (40) | 3 (15) | 4 (16) | 3 (17.6) |
| Smoking history * | | | | |
| Yes | 3 (60) | 2 (10) | 4 (16) | 2 (11.8) |
| No | 2 (40) | 18 (90) | 21 (84) | 15 (88.2) |
| Alcohol history | | | | |
| Yes | 2 (40) | 1 (5) | 3 (12) | 1 (5.9) |
| No | 3 (60) | 19 (95) | 22 (88) | 16 (94.1) |
| Histological grading | | | | |
| Well-differentiated | 2 (40) | 12 (60) | 16 (64) | 11 (64.7) |
| Mod-differentiated | 1 (20) | 7 (35) | 9 (36) | 6 (35.3) |
| Poorly-differentiated | 0 (0) | 0 (0) | 0 (0) | 0 (0) |

| | | | | |
|---|---|---|---|---|
| Unknown | 2 (40) | 1 (5) | 0 (0) | 0 (0) |
| Clinical stage | | | | |
| I/II | 1 (20) | 4 (20) | 6 (24) | 3 (17.6) |
| III/IV | 4 (80) | 16 (80) | 19 (76) | 14 (82.4) |
| Lymph node involvement (Pathologic) | | | | |
| Negative | 2 (40) | 9 (45) | 11 (44) | 7 (41.2) |
| Positive | 2 (40) | 7 (35) | 10 (40) | 7 (41.2) |
| Unknown | 1 (20) | 4 (20) | 4 (16) | 3 (17.6) |
| T classification (Pathologic) | | | | |
| T1/T2 | 4 (80) | 10 (50) | 12 (48) | 9 (52.9) |
| T3/T4a | 1 (20) | 10 (50) | 13 (52) | 8 (47.1) |
| Tumor site | | | | |
| Alveolar gum | 2 (40) | 7 (35) | 13 (52) | 6 (35.3) |
| Tongue | 2 (40) | 4 (20) | 3 (12) | 3 (17.6) |
| Palate | 0 (0) | 4 (20) | 3 (12) | 3 (17.6) |
| Other | 1 (20) | 5 (25) | 6 (24) | 5 (29.4) |
| Recurrence | | | | |
| Locoregional recurrence | 0 (0) | 2 (10) | 3 (12) | 2 (11.8) |
| Lymph node metastasis | 2 (40) | 2 (10) | 2 (8) | 2 (11.8) |
| Distal metastasis | 0 (0) | 2 (10) | 2 (8) | 2 (11.8) |
| No recurrence | 3 (60) | 14 (70) | 18 (72) | 11 (64.7) |

TN paired, tumor-normal matching among participants

The Chi-squared or Fisher's exact testswere used to compare the Microarray Study and Quantitative Reverse Transcription Polymerase Chain Reaction Validation Groups.

$* p < 0.05$

**Table 2. Primer pairs for reverse transcription polymerase chain reaction validation of candidate biomarker gene expression levels.**

| | Gene Symbol | RefSeq_ID | Sequences (5'->3') | |
|---|---|---|---|---|
| Target genes | *ADAM15* | NR_040773.1 | Forward | CAG GAG ACA GCT CCC AGT GT |
| | | | Reverse | CAA AAG CAT TTC CCC GAG TA |
| | *CDC7* | NM_003503.3 | Forward | GCT CAG CAG GAA AGG TGT TC |
| | | | Reverse | AGC TTT TGT GGT GGA CTG CT |
| | *IL12RB2* | NM_001258216.1 | Forward | ACT GGA GCC TCA GCA CAT CT |
| | | | Reverse | AGC CTC ACC ACT CAG AGC AT |
| | *TNFRSF8* | NM_001243.3 | Forward | GCT CAG ATG TTT TGG GGA AA |
| | | | Reverse | AGA CAC CCA CTC CAT CCT TG |
| Housekeeping genes | *β-actin* | NM_001101.3 | Forward | AGA GCT ACG AGC TGC CTG AC |
| | | | Reverse | GGA TGC CAC AGG ACT CCA |
| | *HPRT* | NM_000194.2 | Forward | ATC CAT TGG AGG GCA AGT C |
| | | | Reverse | GAG CTT TTT AAC TGC AGC AAC TT |

*HPRT*, hypoxanthine phosphoribosyltransferase

*ADAM15*, a disintegrin and metalloproteinase metallopeptidase domain 15

*CDC7*, cell division cycle 7 homolog

*IL12RB2*, interleukin 12 receptor, beta 2

*TNFRSF8*, tumor necrosis factor receptor superfamily, member 8

**Table 3. Fifteen candidate genes whose expression differed between normal and tumor tissue in oral squamous cell carcinoma as indicated by microarray analysis (tumor–normal-paired; *n* = 5).**

| Gene Symbol | Gene name /Accession No. | Role | Disease relationship |
|---|---|---|---|
| C1orf93 (*FAM213B*) | family with sequence similarity 213, member B (NM 001195736.1) | Catalyzes the reduction of prostaglandin H2 to prostaglandin F2α | Unknown |
| *ARHGEF16* | Rho guanine nucleotide exchange factor 16 (NM 014448.3) | Role in chemotactic cell migration and mediate activation of CDC42 by HPV16 | Unknown(Oliver *et al*, 2011) |
| *ZBTB48* | zinc finger and BTB domain containing 48(NM 005341.2) | Binds to and regulates the J and/or S elements in MHC II promoter | Unknown |
| *TNFRSF8* | tumor necrosis factor receptor superfamily, member 8 (NM 001243.3) | Positive regulator of apoptosis. Limit the proliferative potential of autoreactive CD8 effector T cells and protect the body against the autoimmunity. | Hodgkin disease, anaplstic large cell lymphoma (Durkop *et al*, 1992) |
| *UTP11L* | UTP11-like, U3 small Nucleolarribonucleoprotein (NM 016037.3) | Involved in nucleolar processing of pre-18S ribosomal RNA | Unknown |

| | | | |
|---|---|---|---|
| *PLK3* | polo-like kinase 3 (NM 004073.2) | Play a role in regulation of cell cycle progression and tumorigenesis | Cancer, tumor (Li *et al*, 1996) |
| *IL12RB2* | interleukin 12 receptor, beta 2 (NM 001258214.1) | Role in the differentiation of T helper 1 cell and promote the proliferation of T cells as well as natural killer cells. | Allergic asthma, Burkitt lymphoma, autoimmunity (Bassuny *et al*, 2003) |
| *WDR63* | WD repeat domain 63 (NM 145172.3) | Unknown | Unknown |
| *CDC7* | cell division cycle 7 homolog (NM 001134419.1) | Phosphorylate critical substrates that regulate the G1/S phase transition and/or DNA replication | Tumor, cancer, ataxia telangiectasia (Tenca *et al*, 2007) |
| *KCNC4* | potassium voltage-gated channel, Shaw-related subfamily, member (NM 001039574.2) | Mediates the voltage-dependent potassium ion permeability of excitable membranes | Cancer, head and neck tumor (Menendez *et al*, 2010) |
| *DCLRE1B* | DNA cross-link repair 1B (NM 022836.3) | Role in telomere maintenance and protection during S-phase | Hoyeraal-Hreidarsson syndrome (Lenain *et al*, 2006) |
| *LCE1D* | Late cornified envelope protein 1D (NM 178352.2) | Precursors of the cornified envelope of the stratum corneum | Eczema (Marenholz *et al*, 2011) |
| *ATP8B2* | ATPase, class I, type 8B, member 2 (NM 001005855.1) | Operate at the plasma membrane and essential for endoplasmic reticulum exit | Unknown (Harris,Arias, 2003) |

| | | | |
|---|---|---|---|
| *ADAM15* | ADAM metallopeptidase domain 15 (NM 001261464.1) | Role in cell-cell interaction, cell migration, signal transduction, wound healing and pathological neovascularization. | Osteoarthritis, cancer, metastasis (Zhong *et al*, 2008) |
| *KLHDC7A* | kelch domain containing protein 7A (NM 152375.2) | Unknown | Diabetic retinopathy, autism (Kuwano *et al*, 2011) |

*CDC42,* cell division control protein 42
*HPV16,* human papillomavirus 16
*MHC,* major histocompatibility complex

Table 4. The correlation between the clinical parameters and the candidate genes determined by the discriminant analysis.

| | *ADAM15* | *CDC7* | *IL12RB2* | *TNFRSF8* | Signature of 4 candidate genes |
|---|---|---|---|---|---|
| **Variables** | | | | | |
| **Primary site** | | | | | |
| Eigen value | 0.186 | 0.232 | 0.125 | 0.302 | 0.658 |
| Wilks' Lambda | 0.843 (p=0.512) | 0.812 (p=0.421) | 0.889 (p=0.662) | 0.768 (p=0.312) | 0.450 (p=0.654) |
| Accuracy of classification | 35.3 % | 47.1 % | 23.5 % | 41.2 % | 64.7 % |
| **Tissue heterogeneity (grade of differentiation)** | | | | | |
| Eigen value | 0.004 | 0.025 | 0.000 | 0.078 | 0.277 |
| Wilks' Lambda | 0.996 (p=0.808) | 0.976 (p=0.550) | 1.000 (p=0.968) | 0.928 (p=0.297) | 0.783 (p=0.529) |
| Accuracy of classification | 35.3 % | 47.1 % | 47.1 % | 58.8 % | 76.5 % |
| **Stage** | | | | | |
| Eigen value | 0.073 | 0.114 | 0.120 | 0.000 | 0.111 |
| Wilks' Lambda | 0.932 (p=0.313) | 0.898 (p=0.212) | 0.893 (p=0.199) | 1.000 (p=0.986) | 0.900 (p=0.849) |
| Accuracy of classification | 52.9 % | 47.1 % | 70.6 % | 35.3 % | 58.8 % |

**Table 5. The diagnostic validity of the candidate genes for OSCC determined by ROC curves and logistic regression analysis.**

| | *ADAM15* | *CDC7* | *IL12RB2* | *TNFRSF8* |
|---|---|---|---|---|
| ROC curve | | | | |
| Sensitivity | 52.9 % | 52.9 % | 35.3 % | 35.3 % |
| Specificity | 82.4 % | 82.4 % | 94.1 % | 88.2 % |
| Youden index | 0.3529 | 0.3529 | 0.2941 | 0.2353 |
| AUC[a] (95% CI[b]) | 0.699 (0.518 - 0.844) | 0.645 (0.463 - 0.801) | 0.561 (0.380 - 0.730) | 0.602 (0.420 - 0.765) |
| PPV[c] | 75.0 % | 75.0 % | 85.7 % | 75.0 % |
| NPV[d] | 63.6 % | 63.6 % | 59.3 % | 57.7 % |
| *z* | 2.189 | 1.483 | 0.573 | 1.014 |
| *p* | 0.0286* | 0.1380 | 0.5663 | 0.3104 |
| Logistic regression analysis | | | | |
| OR[e] | 7.271 | 26.794 | 0.001 | 0.000 |
| *p* | 0.029* | 0.046* | 0.980 | 0.058 |

[a] AUC : area under the curve calculated

[b] CI : confidence interval

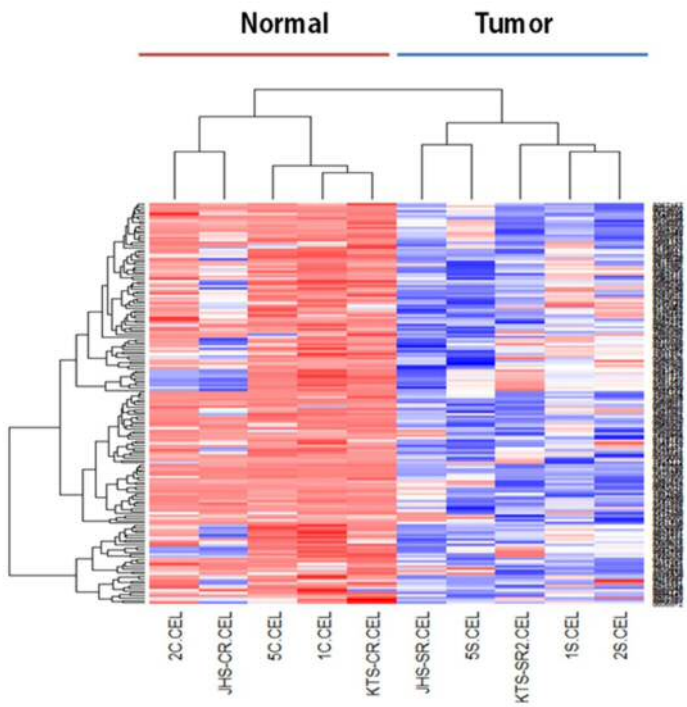
[c] PPV : positive predictive value

[d] NPV : negative predictive value

[e] OR : odds ratio

* $p < 0.05$

## Figures and Legends



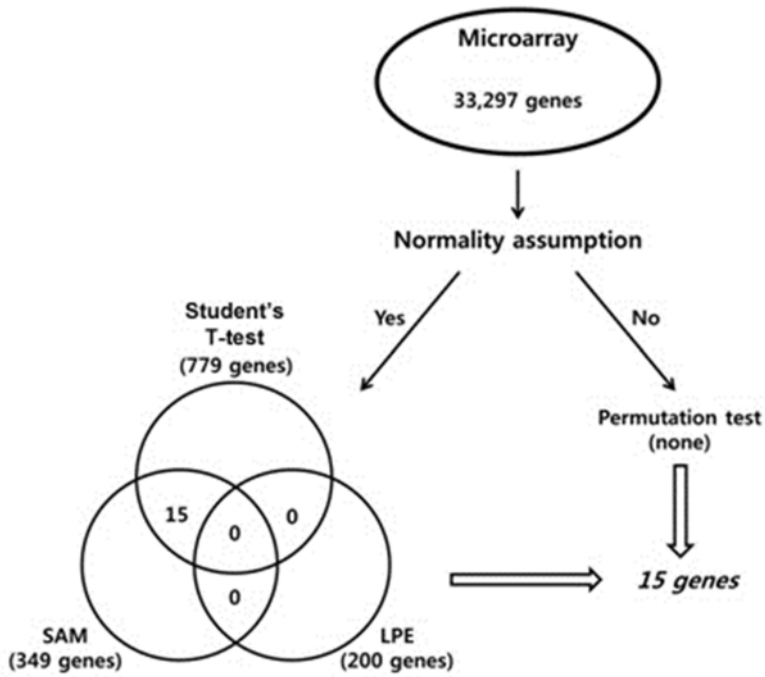



Figure 1. Hierarchical clustering of microarray gene expression data. The dendogram at the bottom lists all samples arrayed and measures their degree of relatedness in terms of gene expression. All samples were coded with numbers, as shown in Table 1. The colored bar beneath the sample identifiers marks samples from patients, where the Normal Group of patients with oral squamous cell carcinoma are in red and those from the Tumor Group are in blue (A). Genes identified by all four statistical tests (the permutation, LPE and t-tests and SAM) were selected as candidate genetic markers. The 15 genes passed Student's t-test and SAM analysis (B).

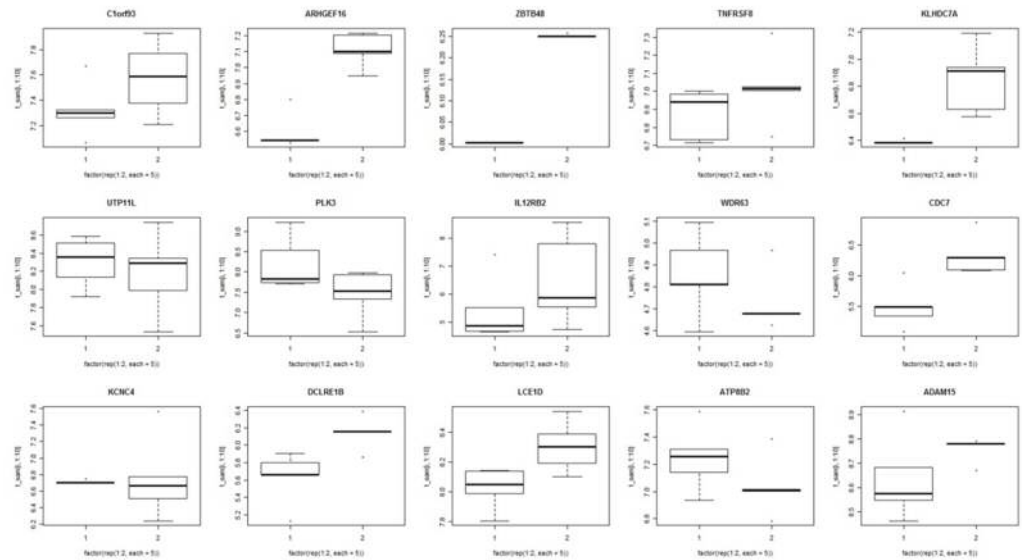



Figure 2. Differential expression of 15 genes selected by microarray in oral squamous cell carcinoma and normal tissue. Five genes were downregulated, while the other 10 were upregulated. The full line corresponds to the median value for each group.

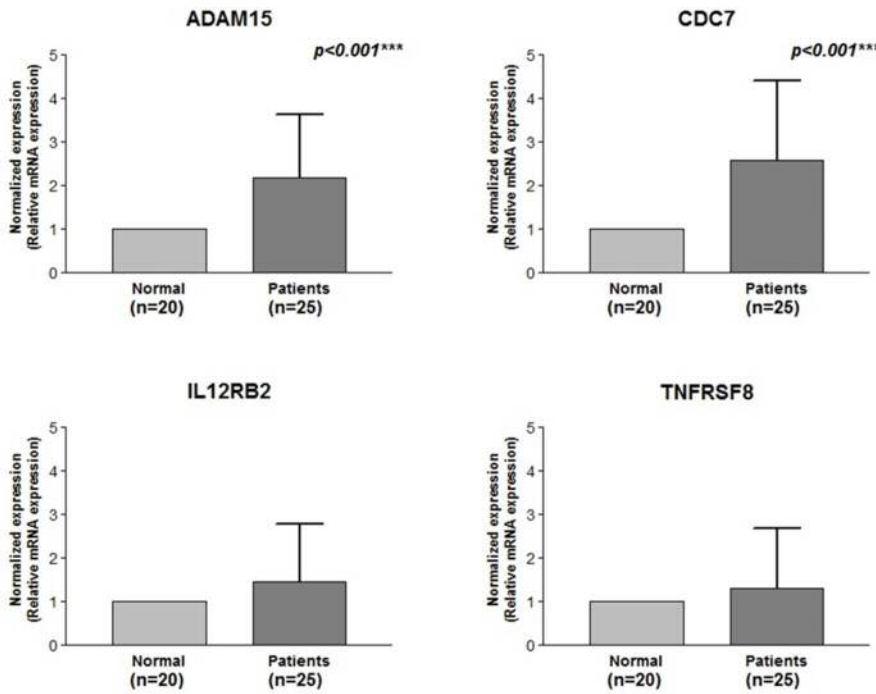



Figure 3. qRT-PCR validation of microarray gene expression data from tested individuals from the Tumor Group. The quantitative reverse transcription polymerase chain reaction tests were performed with primer sets specific to *ADAM15, CDC7, IL12RB2* and *TNFRSF8* in 25 patients with oral squamous cell carcinoma and 20 normal matching samples. p-values (Student's t-test) are presented. (*P < 0.05; **P < 0.01; ***P < 0.001)

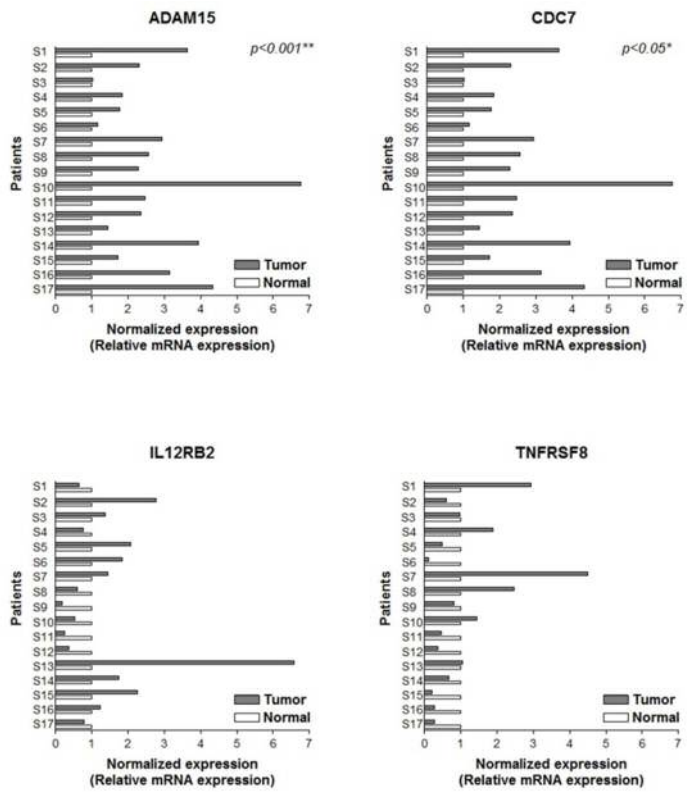



Figure 4. Bar chart representation of the mRNA expression of selected genes in tumor and normal patient-matched specimens (tumor–normal-paired). Normalised expression was determined by 2-ΔΔCt, by comparing threshold cycle values of *ADAM15* with *β-actin*. Overexpression of all genes in tumor tissues is clearly demonstrated. Number of cases analysed and corresponding p-values (Student's t-test) are provided. (*P < 0.05; **P < 0.01; ***P < 0.001)

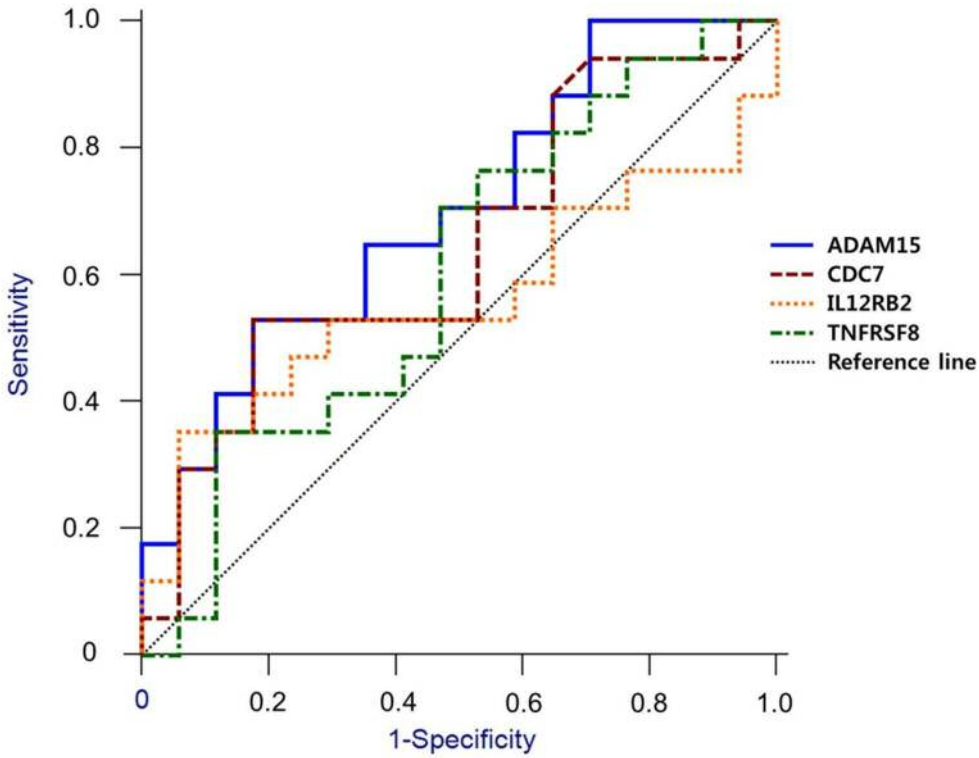



Figure 5. Receiver operating characteristic (ROC) curve for comparing diagnostic validity of the candidate genes. The area under the curve (AUC) of *ADAM15, CDC7, IL12RB2* and *TNFRSF8* were 0.699, 0.645, 0.561 and 0.602, respectively. Among them, only ADAM5 showed statistical significance ($p < 0.05$).

**국문초록**

# 구강편평상피세포암종의 조기 진단을 위한 후보 생체지표의 규명


**김 용 덕**

**서울대학교대학원 치의학과 구강악안면외과학 전공**

**(지도교수 이 종 호)**



**연구목적**

구강편평상피세포암종 (oral squamous cell carcinoma : OSCC)은 빠른 성장 속도와 침습성, 빈번한 림프절의 전이 등 그 예후가 나쁜 편이다. 따라서 OSCC의 조기진단은 매우 중요하며, 이를 위해 임상적으로 적용가능한 생체지표의 확립이 필요하다. 본 연구는 transcriptional signature profiling 을 분석하여 유전학적인 연구와 생물정보학을 이용하여 OSCC의 조기진단에 사용할 수 있는 새로운 생체지표 후보군을 제시하고자 시행하였다.

**연구 방법**

서울대학교 치과병원 구강악안면외과에서 수술 받은 OSCC 환자의 병소와 정상조직을 습득하여 mRNA를 분리하고 cDNA를 합성하였다. 동일 환자의 두 조직에서 유전자발현을  마이크로어레이 분석법으로 상호 비교, 분석하였다. 발현된 유전자 (differentially expressed genes: DEG)는 permutation, local pooled error(LPE), Student's t-tests, significance analysis of microarrays (SAM)을 이용하여 생체지표 후보군으로 선별되었다. 후보군 검증을 위해 통계적으로 유사한 분포를 갖는 validation group의 병소 및 정상조직의 유전자로 정량적역전사중합효소연쇄반응 (qRT-PCR) 분석법으로 적합성을 확인하였다. 임상적 변수와 유전자발현을 비교하였으며, Youden index를 이용하여 후보군 단독 혹은 복합적인 진단적인 가치를 비교, 분석하였다.

**결과**

33,297개의 유전자 중에서 15개의 유전자가 Student's t-test ($p<0.05$)와 SAM을 통해 유의하게 발현됨이 확인되었다. 이들 중 정상조직군에 비해 10

개의 유전자가 과발현되었고 5개의 유전자가 발현이 감소하여, OSCC에서 특징적으로 발현됨을 확인할 수 있었다. 각 유전자의 기능과 존재론적 분석을 기초로 4종 유전자, 즉 *ADAM15, CDC7, IL12RB2, TNFRSF8*를 후보군으로 제시하였으며, qRT-PCR로 검증하였다. 특히 *ADAM15*과 *CDC7*은 모든 환자들에게서 통계적으로 유의성있게 과발현되었다. 임상적 변수와 유전자발현에 관한 상관관계는 후보 유전자 각자보다 복합적인 조합에서 좀더 정확도가 높았으며, 진단적 가치는 ADAM15과 CDC7에서 유의성있게 나타났다.

**결론**

본 연구 결과는 *ADAM15, CDC7, IL12RB2, TNFRSF8* 4종의 유전자가 OSCC의 조기진단을 위한 생체지표 후보군으로서 가능성이 있음을 시사하였고, 향후 OSCC의 생체지표를 연구하는데 중요한 기시점이 될것으로 본다.

---

주요어 : 생체지표, 구강편평상피세포암종, 마이크로어레이, 정량적 역전사 중합효소 연쇄반응, 유전자 발현 분석

학 번 : 2006-30985