Molecular genetic study of novel biomarkers for early diagnosis of oral squamous cell carcinoma

Kim Yong-Deok [1], Jeon Eun-Hyoung [2], Kim Yeon-Sun [2], Pang Kang-Mi [3], Lee Jin-Yong [4], Cho Sung-Hwan [5], Kim Tae-Yun [1], Park Tae-Sung [5], Kim Soung-Min [1], Kim Myung-Jin [1], Lee Jong-Ho [1]

[1] DDS, PhD. ProfessorDepartment of Oral and Maxillofacial Surgery, Graduate School of Dentistry, Seoul National University, Seoul, Republic of Korea
[2] Diagnostic DNA chip center, Institute of Medical Research, Seoul National University, Seoul, Republic of Korea
[3] Department of Oral and Maxillofacial Surgery, Ajou University School of Medicine, Suwon, Republic of Korea
[4] Department of Oral and Maxillofacial Surgery, Korea University Guro Hospital, Seoul, Republic of Korea
[5] Department of Statistics, College of Natural Sciences, Seoul National University, Seoul, Republic of Korea

Correspondence:
Department of Oral and Maxillofacial Surgery
School of Dentistry, Seoul National University
275-1, Yeongeon-dong, Jongno-gu
Seoul, Republic of Korea
leejongh@snu.ac.kr

Abstract

Objectives: Early detection and treatment of an oral squamous cell carcinoma (OSCC) is critical because of its rapid growth, frequent lymph-node metastasis, and poor prognosis. However, no clinically-valuable methods of early diagnosis exist, and genetic analysis of OSCCs has yielded no biomarkers.

Study Design: We investigated the expression of genes associated with inflammation in OSCCs via a quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analysis of microarray data. Tumor and normal tissues from five patients with an OSCC were used for microarray analysis. Differentially-expressed genes, identified using permutation, local pooled error (LPE), t-tests, and significance analysis of microarrays (SAM), were selected as candidate genetic markers.

Results: Two groups corresponding to tissue identity were evident, implying that their differentially-expressed genes represented biological differences between tissues. Fifteen genes were identified using the Student's paired t-test ($p<0.05$) and the SAM, with a false discovery rate of less than 0.02. Based on gene expression, these 15 genes can be used to classify an OSCC. A genetic analysis of functional networks and ontologies, validated by using a qRT-PCR analysis of the tissue samples, identified four genes, ADAM15, CDC7, IL12RB2 and TNFRSF8, that demonstrated excellent concordance with the microarray data.

Conclusions: Our study demonstrated that four genes (ADAM15, CDC7, IL12RB2 and TNFRSF8) had potential as novel biomarkers for the diagnosis and the treatment of an OSCC.

Key words: *Biomarker, microarray, quantitative reverse transcription polymerase chain reaction, oral squamous cell carcinoma, gene expression profiling.*

Introduction

An oral squamous cell carcinoma (OSCC), a subtype of head and neck squamous cell carcinomas, 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 an OSCC exhibit an advanced stage (Stage III or IV) at diagnosis because of its similarity to inflammatory disease, its long asymptomatic period and its challenging clinical differentiation (4,5). Therefore, early detection of an OSCC is critical. 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 an OSCC and variations in the pathologists' experience.

Recently, mRNA biomarkers for an 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 by using the quantitative reverse transcriptase polymerase chain reaction (qRT-PCR), has been used to identify the genes underlying OSCC pathogenesis, which include IL-8 and VEGF (6). Some investigators have employed a multiple-gene model of 25 biomarkers with 86%-89% accuracy (7). 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 (8) or to the limited predictive power of microarray-based models to correlate the clinical endpoint with gene expression (9). In this study, we investigated the expression of genes associated with cancer and inflammation in patients with an OSCC by using microarray analysis, which was validated by using the qRT-PCR. These results were assessed by bioinformatics verification. With this approach, we tried to identify novel biomarkers for OSCC.

Material and Methods

1.1. Patients

TN-paired 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 34 consecutive patients undergoing therapeutic surgical resection for an 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, USA) until use 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.

1.2. 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 by using the RNeasy® Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. To quantify and analyze the integrity of RNA, we used 2100 Bioanalyzer® (Agilent Technologies, USA). We reversely transcribed 10 µg of total RNA in the presence of an oligo (dT) T7 primer (iScript™ Select cDNA Synthesis Kit; Bio-Rad Laboratories, USA). cDNA was used for in-vitro transcription amplification in the presence of biotinylated nucleotides.

1.3. Microarray analysis

Gene expression was analyzed using the Affymetrix preparation protocol before hybridization to a Gene-Chip® 1.1 human genome microarray (Affymetrix, 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) 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) were derived by using the robust multi-array average (RMA) algorithm, as implemented by the Expression Console v1.1.1 (Affymetrix). 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.

1.4. Quantitative reverse transcription polymerase chain reaction

Target and housekeeping gene sequences were retrieved from the GenBank® database (National Center for Biotechnology Information (NCBI), USA), and applied primers were manually designed using the Primer-BLAST® tool (NCBI). 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 (NM_000194.2) and β−actin (NM_001101.3), were used for data normalization, and relative quantification was performed by using a relative standard curve analysis with a CFX Connect™ Real-Time PCR Detection System (Bio-Rad Laboratories) and SYBR® Green I (Bio-Rad Labora-

Table 1. Clinicopathological characteristics of patients in this study and the Qualitative Reverse Transcription Polymerase Chain Reaction Validation Group.

Parameter	Study (Microarray)	Validation (qRT-PCR)		
	TN paired (n=5)	Normal (n=20)	Tumour (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)
Tumour 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, tumour-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.
* $p < 0.05$

tories) detection. cDNA (10ng) was dissolved in iQ™ SYBR® Green Supermix. 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). The fold change of target gene expression in each treated sample, relative to the con-

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

	Gene Symbol	RefSeq_ID	Sequences (5'->3')
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
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

HPRT, hypoxanthine phosphoribosyltransferase
ADAM15, a disintegrin and metalloproteinase metallopeptidase domain 15
CDC7, cell division cycle 7 homolog
IL12RB2, interleukin 12 receptor, beta 2
TNFRSF8, tumour necrosis factor receptor superfamily, member 8

trol sample, was derived from the negative Ct value: Ct (GAPDH)−Ct (target gene).

1.5. Data analysis and statistics

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 by using four statistical tests (permutation test, local pooled error (LPE) (10), Student's paired t-test, and significance analysis of microarrays (SAM) (11), were selected as candidates for genetic biomarkers (12), and their normalities were tested using the Shapiro-Wilk test (13). After the normality had been tested, genes that were normally distributed were assessed using the 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, we tried to use FDR values as close to 0% as possible, and those for the SAM and the permutation test were set to 0%. On the other hand, the FDRs for the LPE and the 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, Ostend, 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).

Results

2.6. Patient characteristics

Thirty-three patients with an OSCC, 17 men (51.5%) and 16 women (48.5%), were included in this study (Table 1). We analyzed the DEGs in five patients by using microarray analyses 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 participants, 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).

2.7. Statistical analysis of candidate genes identified by using a microarray analysis

A t-test was performed to detect the DEGs. However, the test assumes that all data follow a normal distribution. Therefore, a non-parametric permutation test was also conducted to identify DEGs, and its results were evaluated using the SAM and the LPE test. 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 (10). 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 (Fig. 1) (11).

Of the 33,297 genes on the microarray, the numbers of genes with significance, which were assessed with the t-
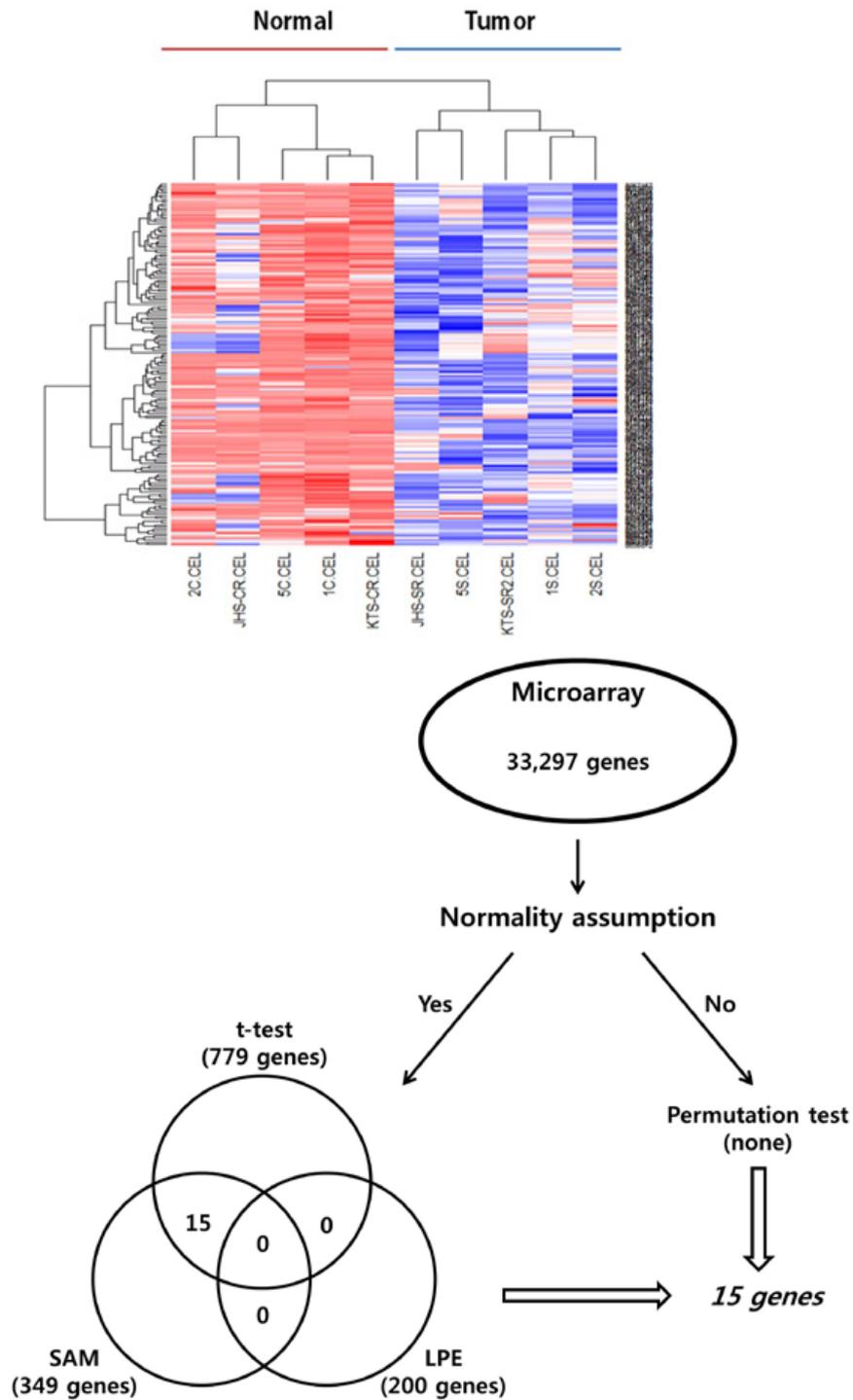
Fig. 1. Hierarchical clustering of microarray gene expression data. The dendogram at the bottom lists all samples arrayed and measures their degree of relatedness in terms of gene expression. All samples were coded with numbers, as shown in table 1. The colored bar beneath the sample identifiers marks samples from patients, where the Normal Group of patients with oral squamous cell carcinoma are in red and those from the Tumour 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 paired t-test and SAM analysis (B).

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 15 genes did in both the t-test and the SAM. However, no genes showed significance in the non-parametric permutation test. The 15 genes showed differential expression patterns; 10 genes were predominantly upregulated and five were predominantly downregulated (Table 3). Four statistical tests (the permutation test, the LPE, the Student's paired t-test, and the SAM) were used to identify the DEGs, and common DEGs identified by using the t-test and the SAM were selected as biomarker candidates.

Fifteen genes passed the Student's paired t-test ($p <$

Table 3. Fifteen candidate genes whose expression differed between normal and tumour tissue in oral squamous cell carcinoma as indicated by microarray analysis (tumour–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	tumour 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 tumourigenesis	Cancer, tumour (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 etal, 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	Tumour, 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 tumour (Menendez et al, 2010)
DCLRE1B	DNA cross-link repair 1B (NM 022836.3)	Role in telomere maintenance and protection during S-phase	Hoyeraal-Hreidarssonsyndrome(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
Note: GeneChip® 1.1 human genome microarrays (Affymetrix, Inc., Santa Clara, CA, USA) were used to identify the difference in RNA expression patterns in tissues from five cancer patients and five matched normal subjects

0.05) and the SAM analysis with an FDR of <0.02. To test the ability of this signature gene set to classify the OSCC and the Normal groups, we performed an average linkage hierarchical clustering analysis. Based on gene expression, this 15-gene set has classification power for an OSCC (Fig. 2). In this set, upregulated genes were FAM213B, ARHGEF16, ZBTB48, DCLRE1B, LCE1D, KLHDC7A, TNFRSF8, IL12RB2, CDC7 and ADAM15 while UTP11L, PLK3, WDR63, KCNC4 and ATP8B2 were downregulated.

2.8. 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 signalling and movement (ARGHEF16, KCNC4 and ADAM15), transcription (UTP11L and ATP8B2), cell division (PLK3, CDC7 and DCLRE1B) and keratinization

(LCE1D) (14-24). After a literature review, we excluded well-known genes associated with carcinogenesis, such as PLK 3 and KCNC 4, and unidentified genes, including WDR63 and KLHDC7A. Among the remaining genes, upregulated ones were chosen as candidates. Of them, ARHGEF16 and DCLRE1B were excluded due to their unknown correlation with the tumor. Finally, IL12RB2, TNFRSF8, ADAM15 and CDC7 remained. According to previous research, these 4 genes are known to be related to inflammation, carcinogenesis or metastasis (14,16,18,19).

2.9. Quantitative reverse transcriptase polymerase chain reaction validation of gene expression

Expressions of the four candidate genes (ADAM15, CDC7, IL12RB2 and TNFRSF8) were assayed by using the qRT-PCR in 28 OSCC samples to correlate microarray data with the qRT-PCR data and to explore potential biomarker expression (Fig. 3). Relative expression levels of the four genes were obtained by normalization to
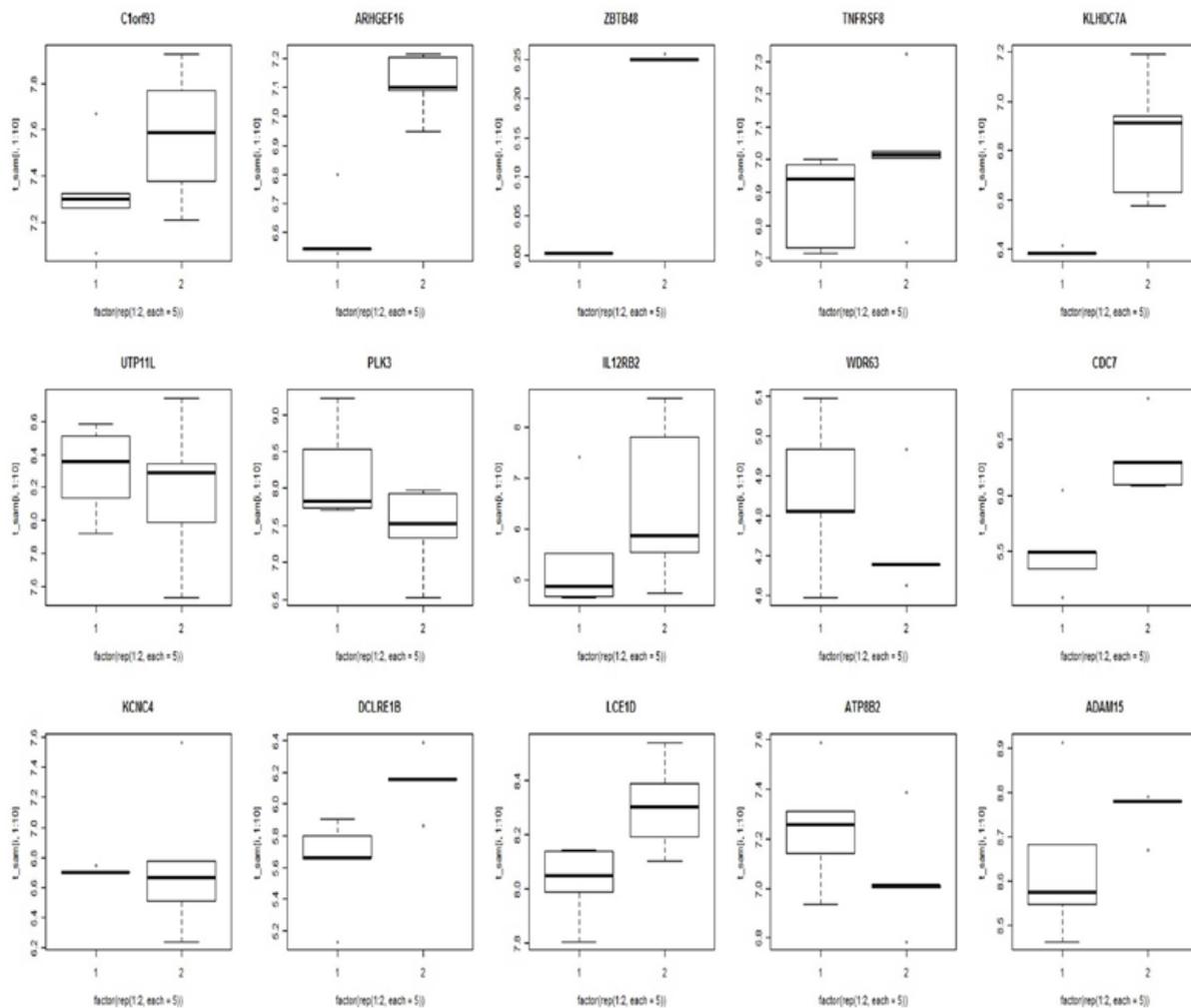
Fig. 2. Differential expression of 15 genes selected by microarray in oral squamous cell carcinoma and normal tissue. Five genes were downregulated, while the other 10 were upregulated. The full line corresponds to the median value for each group.
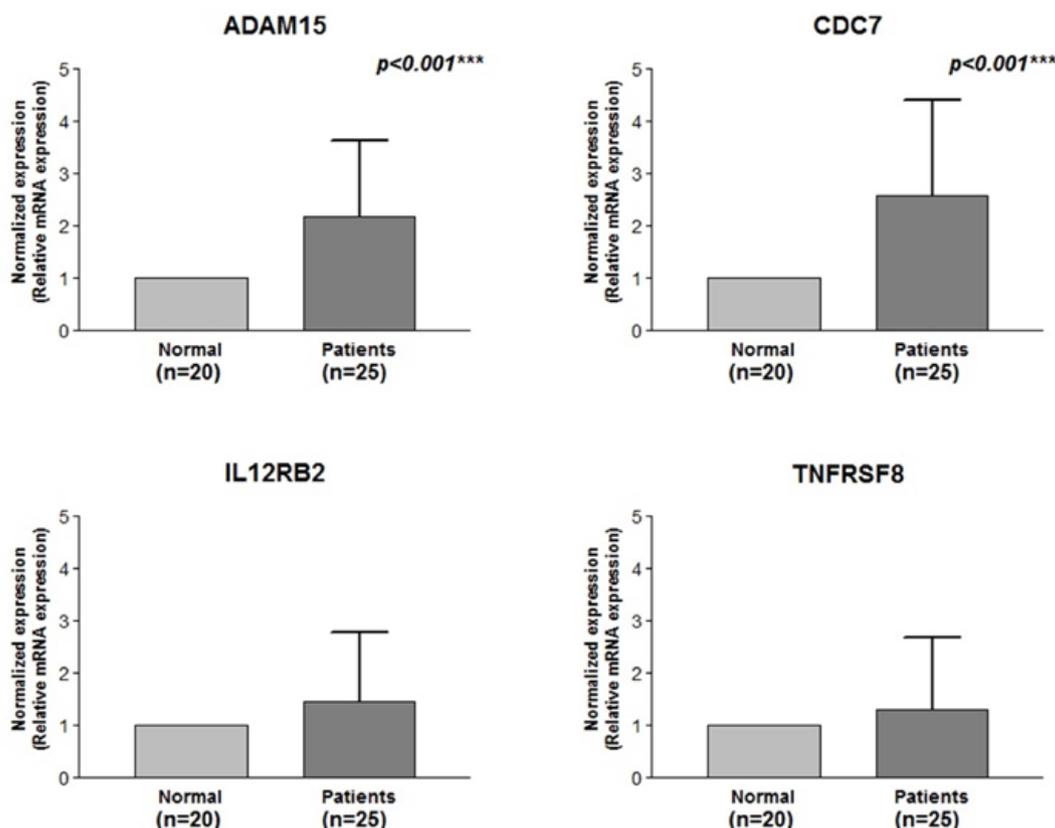
Fig. 3. Quantitative reverse transcription polymerase chain reaction validation of microarray gene expression data from tested individuals from the Tumour Group. The quantitative reverse transcription polymerase chain reaction tests were performed with primer sets specific to ADAM15, CDC7, IL12RB2 and 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)

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 by using the qRT-PCR ($p<0.001$ and $p<0.001$, respectively), thus showing excellent agreement 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.

We also measured the mRNA expressions of the selected genes in 17 TN-paired tissue samples by using the qRT-PCR. The normalized expression was calculated by using the 2-ΔΔCt method, as previously described, to compare normal and tumor tissues. ADAM15 expression was significantly increased in tumor tissue ($p<0.001$) by an average of 2.7-fold compared with normal tissue (Fig. 4). Consistent with these findings, upregulation of CDC7 was evident in tumor tissue ($p < 0.05$). Furthermore, the candidate genes were overexpressed regardless of variability in the degree of expression between samples in the TN-paired group.

2.10. 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 site by using all four candidate genes as signatures resulted in higher accuracy (64.7%) than using the ADAM15, CDC7, IL12RB2 and TNFRSF8 genes by themselves (35.3%, 47.1%, 23.5% and 41.2%, respectively). Similar tendencies were 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 by using each candidate gene alone.

ROC plots are presented in fig. 5, and their diagnostic indices are 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 statistical significance ($p < 0.05$). Similarly, the odds ratio of ADAM15 (7.271; $p = 0.029$) and CDC7 (26.794; $p = 0.046$) were statistically significant.
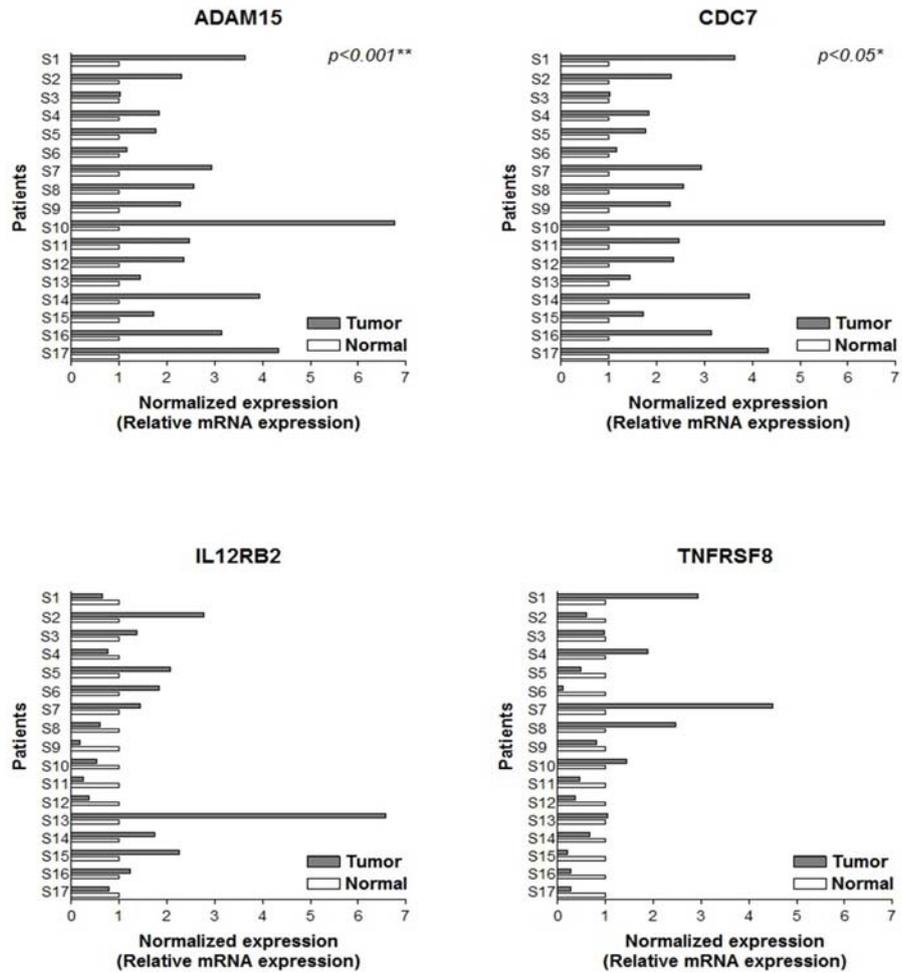
Fig, 4. Bar chart representation of the mRNA expression of selected genes in tumour and normal patient-matched specimens (tumour–normal-paired). Normalised expression was determined by 2-ΔΔCt, by comparing threshold cycle values of ADAM15 with β-actin. Overexpression of all genes in tumour tissues is clearly demonstrated. Number of cases analysed and corresponding p-values (paired t-test) are provided. (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$)

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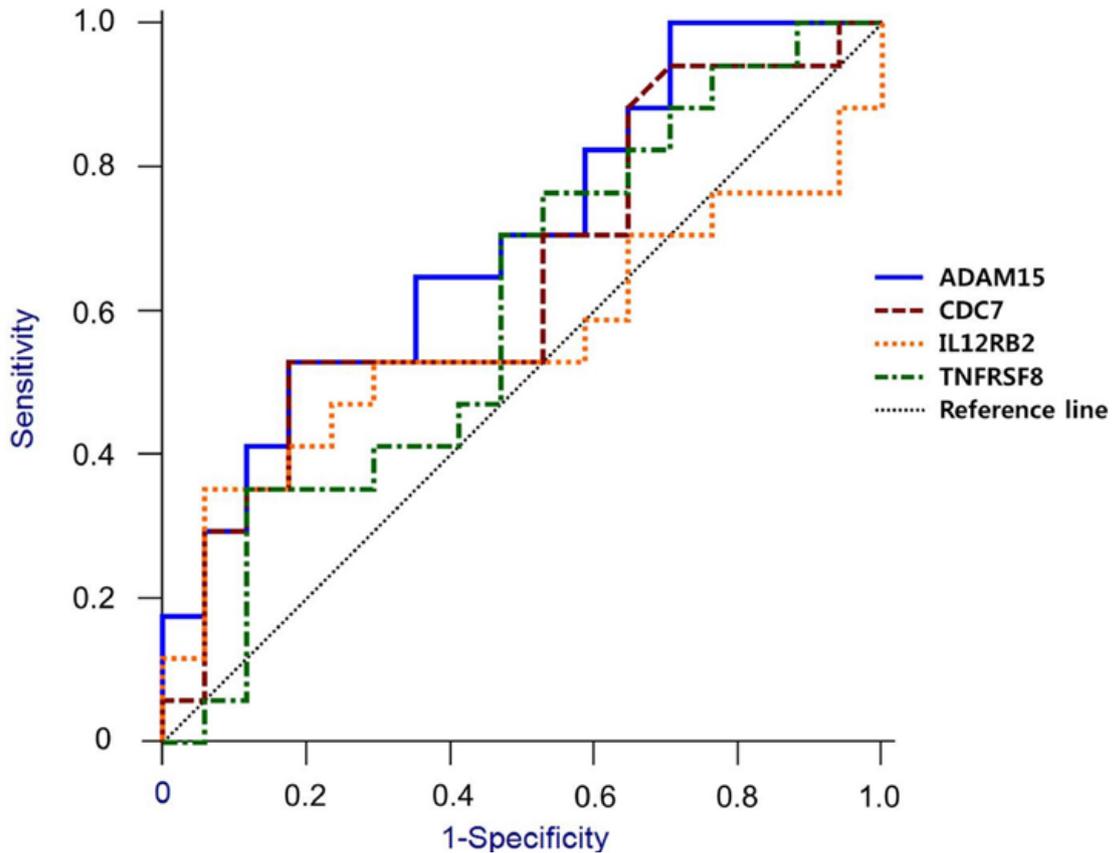
Fig. 5. Receiver operating characteristic (ROC) curve for comparing diagnostic validity of the candidate genes. The area under the curve (AUC) of ADAM15, CDC7, IL12RB2 and TNFRSF8 were 0.699, 0.645, 0.561 and 0.602, respectively. Among them, only ADAM5 showed statistical significance ($p < 0.05$).

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
* $p < 0.05$

Discussion

The five-year survival rate of patients with an 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 (25). Local recurrence and distant metastasis may develop despite an apparently complete excision and histopathologically tumor-free margins, supporting the theory of field cancerization (26). Novel biomarkers could allow a more sensitive and accurate surgical margin to be determined. However, genome research has been hindered by individual and ethnic differences (27).

Although the sample size of the Microarray Study group (n = 5) seems too small, we tried to ensure accurate results in three ways: (1) several statistical analyses (the permutation test, the LPE and the SAM), which can identify significant DEGs with a small number of microarrays, were applied (10), (2) genes carefully selected by using these statistical tools were validated via the qRT-PCR with sufficient samples, and (3) 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 (28), gene expression between smoking and non-smoking subjects did not differ in this study. Further, Méndez et al. observed no statistically-significant differences in gene expression between patients with different histopathologic stages when using oligonucleotide arrays (29). Thus, the effect of smoking history and histopathological type on our results can be expected to be minimal.

To date, many studies have attempted to identify novel biomarkers for an OSCC (3,6,30). 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 (30), but these can often be misleading, as different error variances exist under different biological conditions and in different microarray expression intensity ranges (10). Second, most previous studies mainly focused on well-known genes, such as matrix metalloproteinases (MMPs), p53 and VEGF (6,8). Such an approach could result in a bias against unknown genes. In this study, we tried to overcome possible bias by using a microarray analysis (a relatively unbiased high-throughput method) and to reduce the statistical error by using four kinds of biostatistics verification.

From our results, 15 genes were initially screened (Table 3). As previously mentioned, genes with unknown functions were excluded, and tumor-related ones (ARHGEF16, DCLRE1B, TNFRSF8, PLK3, IL12RB2, CDC7, KCNC4 and ADAM15) were selected. After that, upregulated genes (ARHGEF16, DCLRE1B, TNFRSF8, IL12RB2, CDC7 and ADAM15) were selected according to hypoth-

esis that upregulated biomarkers are more valuable because their overexpression will increase the sensitivity of diagnosis (31). Among them, 4 genes (ADAM15, CDC7, IL12RB2 and TNFRSF8) were selected for further evaluation because of their involvement in malignancy.

ADAM15 is the only member of the disintegrin and metalloproteinase (ADAM) family and is involved in T-cell-mediated immune responses and pathologic neovascularization (32). It is upregulated in breast, stomach, lung, pancreas, and prostate cancer and in metastatic progression (33). Conversely, its tumor suppression in melanomas and colon carcinomas was also proposed (34,35). This inconsistency might result from the complexity of the ADAM15-related signal transduction pathways. In this study, ADAM15 was significantly upregulated in patients with an OSCC.

CDC7, a serine threonine kinase, is a core component of the initiation machinery of DNA replication, and its inhibition in cancer cells induces p53-independent apoptosis (36). Clinically, in many kinds of cancers, CDC7 upregulation is related to tumor anaplasia, aneuploidy, advanced disease stage and lower relapse-free survival (37-39). These results agreed with our data, which showed overexpression of CDC7 in patients with an OSCC.

IL12RB2 encodes a type-1 transmembrane protein, a subunit of the Interleukin 12 (IL-12) receptor complex. IL12RB2 behaves as a tumor suppressor in human chronic B-cell malignancies, and its silencing is an early event in B-cell malignant transformation (27,40). Prolonged disease-free survival has also been reported to have been found in colorectal cancer patients with higher expressions of IL12RB2 (41). However, the correlation between IL12RB2 expression and OSCC development is unclear. 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 (42,43), IL12RB2 downregulation might be involved with an OSCC. However, despite overexpression of IL12RB2 in patients with an OSCC in the Microarray Analysis group, no significant differences of IL12RB2 expression were found in our qRT-PCR data.

TNFRSF8 is known to function in apoptosis by activating the caspase pathway (44) and protecting the body against autoimmunity. It is a conventional biomarker of hematologic malignancies (45), 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 (46). Though TNFRSF8 in patients with an OSCC was upregulated in the microarray analysis, qRT-PCR validation showed no significant differences. Hence, further investigation is needed.

Among the four candidates, ADAM15, 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, our results support that inflammation might be correlated with OSCC development, as reported in some recent studies (6,43).

Table 4 presents the correlation between the candidate genes and the clinical parameters (primary site, tissue heterogeneity and stage). 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 (47). The Youden index is widely used to determine the diagnostic threshold, and it is regarded as the most appropriate method of combining biomarkers (48). 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 statistically-significant correlations to an OSCC ($p < 0.05$).

In conclusion, four genes (TNFRSF8, IL12RB2, CDC7 and ADAM15) were presented as candidates for OSCC biomarkers by using a microarray analysis and a qRT-PCR validation with biostatistical verification. Further investigations are needed in order to construct more accurate diagnostic models. To our knowledge, this is the first report on OSCC biomarkers in Korean patients.

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. Sandler HC. Cytological screening for early mouth cancer. Cancer. 1962;15:1119-24.
5. Al-Sarraf M. Chemotherapeutic management of head and neck cancer. Cancer Metastasis Rev. 1987;6:181-98.
6. 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.
7. 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.
8. Reis PP, Waldron L, Perez-Ordonez B, Pintilie M, Galloni NN, Xuan Y, et al. A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. BMC Cancer. 2011;11:437.
9. 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 Cancer. Genes Cancer. 2011;2:914-22.
10. 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.
11. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci. 2001;98:5116-21.
12. Lockstone HE. Exon array data analysis using Affymetrix power tools and R statistical software. Brief Bioinform. 2011;12:634-44.
13. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52:591-611.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.

29. 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.

30. 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.

31. 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.

32. Lucas N, Najy AJ, Day ML. The therapeutic potential of ADAM15. Current pharmaceutical design. 2009;15:2311-8.

33. Duffy MJ, McKiernan E, O'Donovan N, McGowan PM. Role of ADAMs in cancer formation and progression. Clin Cancer Res. 2009;15:1140-4.

34. 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.

35. 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.

36. 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.

37. Kulkarni AA, Kingsbury SR, Tudzarova S, Hong HK, Loddo M, Rashid M, et al. Cdc7 kinase is a predictor of survival and a novel therapeutic target in epithelial ovarian carcinoma. Clin Cancer Res. 2009;15:2417-25.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. Kang MR, Kang JS, Yang JW, Kim BG, Kim JA, Jo YN, et al. Gene expression profiling of KBH-A42, a novel histone deacetylase inhibitor, in human leukemia and bladder cancer cell lines. Oncol Lett. 2012;3:113-8.

45. 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.

46. 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.

47. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3:32-5.

48. Yin J, Tian L. Optimal linear combinations of multiple diagnostic biomarkers based on Youden index. Stat Med. 2014;33:1426-40.