SPANDIDOS PUBLICATIONS

# Gene expression profile of oral squamous cell carcinomas from Sri Lankan betel quid users

MAI LILL SUHR[1], BJARTE DYSVIK[2], OVE BRULAND[3], SAMAN WARNAKULASURIYA[4], ASOKA N. AMARATUNGA[5], INGE JONASSEN[2,6], ENDRE N. VASSTRAND[7] and SALAH O. IBRAHIM[1,7]

[1]Department of Biomedicine, University of Bergen, Jonas Lies Vei 91, N5009 Bergen; [2]Department of Informatics, University of Bergen, HIB, N5020 Bergen; [3]Center of Medical Genetics and Molecular Medicine, Haukeland University Hospital, N5021 Bergen, Norway; [4]Department of Oral Medicine and Pathology, Guy's, King's and St Thomas' School of Dentistry, WHO Collaborating Centre for Oral Cancer and Precancer, King's College London, Denmark Hill Campus, Caldecot Road, London SE5 9RW, UK; [5]Faculty of Dental Sciences, University of Peradeniya, Peradeniya 20400, Sri Lanka; [6]Computational Biology Unit, BCCS, HIB, N5020 Bergen; [7]Dental Faculty-Periodontology, University of Bergen, Årstadveien 17, N5009 Bergen, Norway

Received May 22, 2007; Accepted July 2, 2007

**Abstract.** Oral squamous cell carcinoma (OSCC) is one of the major health problems in Sri Lanka and the disease is associated with the habit of Betel Quid (BQ) chewing. Using 35k oligo microarrays, we analyzed the gene expression profile of 15 Sri Lankan patients diagnosed with OSCCs and pair-wised normal controls and correlated the findings with the clinicopathological data. Following the recording of the scanned array images and data analysis, results for selected candidate genes were verified using QRT-PCR. Upon analysis, a total of 263 genes [71 (27%) of unknown functions previously not reported in OSCCs and 192 (73%) of known functions] were found as differentially expressed between tumors and controls. For the genes with known functions, 66 (34%; such as *COL4A1*, *MMP1*, *MMP3*, *PLAU*, *SPARC* and *KRT19*) were previously reported in OSCC and for the remaining 126 (66%; such as *CD47*, *APOL3*, *RRAGC*, *BPIL1* and *AZGP1*) this is the first report in OSCCs. Hierarchical clustering of the differentially expressed 263 genes grouped the samples into several clusters with the larger one being dominated by tumors of stage 3 and 4. Two cases (a verrucous SCC and an advanced SCC), did not cluster with any of the other samples. We found two main biological pathways (cell communication and integrin-mediated cell adhesion) and 5 gene ontology categories (transcription regulator activity, structural molecule activity, intracellular signaling, cytoskeleton and signal transduction) of relevance to the OSCCs examined. Results from the QRT-PCR verified the results from the microarray experiment. This study provides valuable information on gene expression profile of OSCCs of habitual users of BQ from Sri Lanka. Of particular interest were the list of genes of known and unknown functions and the two biological pathways that we suggest as candidate genes in oral cancers associated with BQ chewing in Southeast Asia, in particular Sri Lanka. The suggested candidate genes might be used as molecular biomarkers in the early detection of the alarming problem of OSCCs in Southeast Asia in association with BQ use. These findings provide valuable information that might help in the selection of possible biomarkers that can be used in early detection of the alarming problem of oral cancer in Southeast Asia.

*Correspondence to*: Dr Mai Lill Suhr, Department of Biomedicine, University of Bergen, Jonas Lies Vei 91, 5009 Bergen, Norway
E-mail: mai.suhr@biomed.uib.no

*Key words:* oral squamous cell carcinoma, microarrays, genomics, betel quid, Sri Lanka

## Introduction

Of the 500,000 new cases of oral squamous cell carcinomas (OSCCs) reported annually, 62% occur in developing countries with the highest rates reported in Sri Lanka and other parts of Southeast Asia (1). In this part of the globe, an estimated 600 million people are reported to chew areca nut and betel quid (BQ) (2). In Sri Lanka, OSCC is linked to the habit of BQ use (3). Areca nut (Areca catechu), the major component of BQ, has been found to contain several alkaloids that give rise to nitrosamines, some of which are carcinogenic (4). It has been suggested that BQ-specific nitrosamines (BQ-SNAs) may act as an adjunct to tobacco-specific nitrosamines (TSNAs), implicated as an etiologic factor for OSCCs (5). BQ-SNAs includes N-nitroso-guvacoline (NG) and 3-(methyl-N-nitrosamino) propionitrile (MNPN) that is also found in tobacco (6). In Sri Lanka, fresh areca nut, slaked lime from seashells, betel leaf and dried (or processed) tobacco are chewed, with the quid placed between the cheek mucosa, the lower gingival and sulcus and sometimes retained during

sleep (2). Buccal mucosa represents the primary site for OSCC among betel quid chewers (2), contrary to tongue cancer, which represents the primary site of cancers in Western countries where cigarette smoking and heavy alcohol consumption are the main causative factors (7). Molecular mechanisms and steps of carcinogenesis in betel quid induced OSCC may differ from those caused by cigarette smoking and alcohol use (8).

Development of OSCC is a multi-step process involving genes related to cell cycle, growth control, apoptosis, DNA damage response and other cellular regulators (9). Understanding the genetic processes and biological pathways involved in the development of OSCC might lead to valuable information that might improve disease classification, early detection and diagnosis, as well as therapeutic planning and drug development (10,11). Microarrays represent a promising tool that makes it possible to explore the expression profile of thousands of genes simultaneously, at the RNA level (12,13). In the literature, there are several microarray studies on OSCCs with promising findings (12-14). Although the influence of life-style factors such as tobacco, alcohol use and nutrition are important to consider in the causation of OSCCs, there is a wide inter-individual difference(s) in susceptibility to chemical carcinogens (15). Since BQ is associated with OSCC development, we applied the 35k human oligo-microarrays (produced and supplied by the Norwegian Microarray Consortium, www.mikromatrise.no) to examine the gene expression profile in 15 cases of OSCCs from Sri Lanka with their corresponding pair-wise normal controls and correlated the findings to patient's clinicopathological parameters.

## Materials and methods

*Patients*. Primary samples (n=15) of OSCCs and their corresponding pair-wised normal controls were acquired from consecutive patients (average age 58.8 years; range 43 to 70; SD ± 8.68) with previously untreated OSCCs operated on at the Department of Maxillofacial Surgery, University Dental School at Peradeniya, Sri Lanka. The Ethics committee at the University of Peradeniya approved the study and written consent was obtained from the participating subjects. Normal control samples were obtained either from the contralateral side of the mouth or from an area that was at least 4 to 5 cm away from the cancerous tissue and was macroscopically normal. After surgery, tissue samples (malignant and normal) were immediately submerged in the tissue storage and RNA stabilization solution, RNAlater™ (Ambion, Inc., Woodlands, TX, USA) and dispatched to the Department of Biomedicine at the University of Bergen, where they were stored at -20˚C until RNA purification and microarray experiments.

All tumors were staged following the 1987 UICC staging system and had their histopathologic diagnosis confirmed by two of the authors (SW/SOI) using either fresh frozen/or 10% formalin-fixed, paraffin-embedded tissue sections stained with haematoxylin and eosin (H&E). The tumors were histologically graded into high, moderate or poorly differentiated carcinomas (16). To rule out gene expression alterations because of stromal cell contamination, we confirmed pathologically that each tumor specimen contained ≥70% cancer cells by analysis of the corresponding H&E-stained sections. For all the patients, data on clinicopathological parameters, such as information on betel quid chewing, were available (Table I).

*Tissue samples and RNA exctraction*. Total RNA was extracted from both tumor and normal controls using TRIzol® reagent (Gibco BRL, Carlsbad, CA, USA)/RNeasy Fibrous Tissue kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's instructions. Quality and quantity of the RNA were determined spectrophotometrically with a Beckman DU®530 Life Science Spectrophotometer (Beckman Coulter, Inc., Fullerton, CA, USA) and by an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).

*cDNA synthesis, hybridization and scanning*. Synthesis and labeling of the cDNA was carried out using Fairplay Microarray Labelling kit (Stratagene, La Jolla, CA, USA), following the manufacturer's instructions. Synthesized cDNA was labeled with Cy™3 (normal cDNA) and Cy™5 (tumor cDNA) monoreactive dyes (Amersham Biosciences, GE Health Care) and samples were hybridized to the human oligonucleotide microarrays containing 34,580 oligonucleotid probes (The Human Genome Oligo Set Version 3.0; Operon Biotechnologies Inc., Huntsville, AL, USA) representing human genes and gene transcripts, printed on Corning Ultra GAPS slides at the Norwegian Microarray Consortium (www.mikromatrise.no). Labeled cDNA was hybridized on the Ventana Discovery® XT System (Ventana Medical Systems Inc., Tucson, AZ, USA) according to the manufacturer's instructions. Slides were scanned by Agilent DNA Microarray Scanner BA (Agilent Technologies, Palo Alto, CA, USA) and the microarray data were stored as tiff format images. The images were further analyzed with GenePix Pro v5.0 (Molecular Devices Corp., Sunnyvale, CA, USA) where bad spots and spots not found were flagged and the final results containing all statistical values were stored as a gpr-file.

*Statistical analysis*. Image quantization files obtained from GenePix Pro were processed and merged into a gene expression matrix using J-Express Pro software package (version 2.6; www.molmine.com) (17). Each array was first pre-processed separately by performing the following steps: i) Spots flagged by Genepix ('bad', 'absent' or 'not found', -100, -75 and -50, respectively) were filtered; ii) and in order to avoid extreme ratios in spots where only one of the channels had a significant signal, a flooring step was applied where intensity values below 30 was set to 30, thereby eliminating unwanted high ratios for spots with intensity near zero; iii) a global lowess normalization was applied to all values left after the filtering step. Thereafter, all in-array replicate spots were merged by a median statistics and inserted into a gene expression matrix where each row corresponds to a gene, each column corresponds to a patient and each cell represents the log (base 2) ratio value for the tumor versus the normal control for one gene in one patient. Since the expression matrix will contain cells with no values (missing values), genes for which more than half of the patients (arrays) had no value were removed. Missing values were set to zero, thus avoiding their contribution in indicating up- or downregulation status of a gene in tumor versus normal control. To prepare the

'linicopathological data of the 15 patients included in the study.

Patients' clinicopathological data

| Sample no | Tumor site | TNM | Stage | Clinical appearance | BQ | Smoking | Alcohol |
|---|---|---|---|---|---|---|---|
| 1 | Alveolus | T2 N0 M0 | 2 | Verrucous | ++ | ++ | ++ |
| 2 | Buccal | T1 N0 M0 | 1 | Exophytic | ++ | - | - |
| 3 | Alveolus | T2 N1 M0 | 3 | Growth | ++ | - | - |
| 4 | Floor | T2 N1 M0 | 3 | Growth | ++ | ++ | ++ |
| 5 | Buccal | T4 N1 M0 | 4 | Growth | ++ | - | - |
| 6 | Alveolus | T4 N1 M0 | 4 | Growth | ++ | - | - |
| 7 | Alveolus | T2 N1 M0 | 3 | Growth | ++ | ++ | ++ |
| 8 | Buccal | T4 N1 M0 | 4 | Exophytic/ulcer | ++ | ++ | - |
| 9 | Lateral tongue | T2 N0 M0 | 2 | Growth | - | ++ | ++ |
| 10 | Buccal | T2 N0 M0 | 2 | White/red ulcer | - | + | ++ |
| 11 | Buccal | T4 N1 M1 | 4 | Growth | ++ | ++ | ++ |
| 12 | Alveolus | T4 N0 M0 | 4 | Growth | ++ | - | - |
| 13 | Buccal | T2 N0 M0 | 2 | Verrucous | ++ | ++ | + |
| 14 | Retromolar | T4 N0 M0 | 4 | Ulcer | ++ | - | - |
| 15 | Buccal | T2 N0 M0 | 2 | White/ulcer | ++ | ++ | + |

TNM, tumor grading Broder's classification system; ++, regular (daily); +, occasionally; -, non-user.

expression matrix for array comparison, we applied scale normalization to reduce differences in expression spread. For finding a gene with a significant difference between tumor and normal control, the relative difference in gene expression d (i)=M(i)/(SE(i)+s) was used where M(i) is the mean log ratio for gene i, SE(i) is the standard deviation of the gene's log ratios and s is an added constant for all genes. In words, d(i) is a student's t-statistics with a fudge factor s which corrects for under-estimated variances resulting in a higher weight to high average fold change versus low variance that can be justified by the noisy nature of microarray experiments. In our case and as suggested by Efron *et al* (18), a0 is set to a 90 percentile of all gene standard errors SE(i).

Since all tumors were labeled with Cy5 and the corresponding normal controls were labeled with Cy3, we used as a control an additional set of hybridizations of 5 pairs of experiments that included primary keratinocytes, dysplastic oral epithelial cell line, OSCC cell line (SCC-25) and two metastatic OSCC cell lines (OSC-2 and G6) that were hybridized twice with a dye swap experiment using the same arrays and identical experimental protocols. This was done to find out whether genes found as differentially expressed were due to dye swap effect or were due to the disease status. In our case, we hypothesized that a gene-specific dye effect would give genes with high s-scores in this matrix since a bias for one gene will give a higher signal with one of the dyes and will have this as a result. Therefore, we have chosen a very low threshold of 0.5 for the s-score, which has resulted in a list of 1276 genes with a possible dye effect.

For the tumor expression matrix, we selected genes with an s-score above 1.0 and obtained 461 genes as differentially expressed either between the tumors and normal controls or alternatively, as a result of the dye effect. To remove genes affected by a dye effect, we removed all genes for which array vendor had reported a possible dye effect and also genes with an s-score above 0.5 in the dye swap expression matrix, resulting in 263 genes that we believe are differentially expressed between tumors and normal controls. We further performed a permutation experiment and generated 1,000 permuted matrices. Each permuted matrix contains the 12034 genes in the original unfiltered data set and the permutation was performed by flipping the sign of each column with a probability of 0.5. Averaged over the 1,000 permuted matrices, our analysis generated 0.47 genes per matrix with a d-score above 1.0 resulting in a false discovery rate of 0.001.

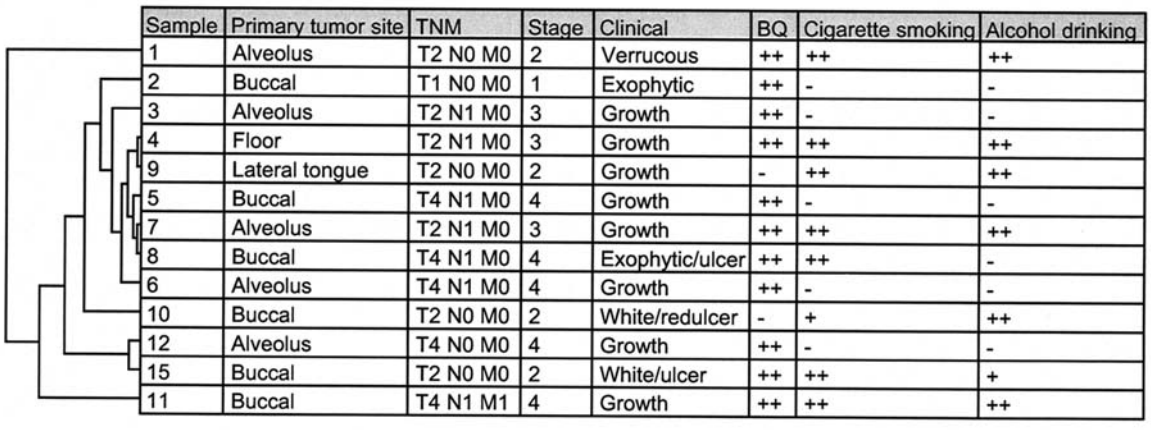
To search for changed biological systems, mapping of the selected genes was first done to a Gene Ontology (GO) Directed Acyclic Graph (DAG) and thereafter comparison of the number of the selected genes (263) in GO terms was done to the number of the genes in a GO DAG based on all genes left from low-level data preparation (12034 genes). We used the Fisher-Irwin exact test to calculate a p-value for all GO terms using a p-value cut-off at 0.05. All terms with less than three selected genes and enrichment score (ratio of the relative frequency of genes from a GO-term in the selected set to the relative frequency of genes from the same term in the full set) below 2 were removed. We searched for genes related to the same biological pathway by performing a KEGG (19-21) (Kyoto Encyclopedia of Genes and Genomes, www.KEGG.com) analysis. We also used the Fisher-Irwin exact test to find statistically significant pathways.

Hierarchical clustering, based on Pearson correlation and average-linkage (WPGMA), was performed to cluster patients with similar gene expression profiles.

*Quantitative real-time RT-PCR*. To validate gene expression profile for selected candidate genes, real-time quantitative RT-PCR was performed for nine genes: *GJA1*, *MMP1*,

Table II. List of the probes for the genes analyzed by QRT-PCR.

| QRT-PCR probes | | |
|---|---|---|
| Gene | Assay ID | Reporter sequence |
| MMP1 | Hs00233958_m1 | TAAAGACAGATTCTACATGCGCACA |
| COL4A1 | Hs00266237_m1 | CCTCCAGGCCTCCCTGTACCTGGGC |
| GJA1 | Hs00748445_s1 | GACCAGTGGTGCGCTGAGCCCTGC |
| HAS3 | Hs00193436_m1 | TGGACTACATCCAGGTGTGCGACTC |
| ANGPTL4 | Hs00211522_m1 | CTGCACCGGCTGCCCAGGGATTGCC |
| CD47 | Hs00179953_m1 | GGCGTGCTGCGGATCAGCTCAGCTA |
| BPIL1 | Hs00257918_m1 | AGCATTGAGCTACGTGTCTGAAATT |
| WFDC2 | Hs00196109_m1 | CACCCTAGTCTCAGGCACAGGAGCA |
| S100A1 | Hs00196704_m1 | CCAGGCCAACCGTGCACTGCTGCAA |
| ACTB | Hs99999903_m1 | GCCTCGCCTTTGCCGATCCGCCGCC |



| Sample | Primary tumor site | TNM | Stage | Clinical | BQ | Cigarette smoking | Alcohol drinking |
|---|---|---|---|---|---|---|---|
| 1 | Alveolus | T2 N0 M0 | 2 | Verrucous | ++ | ++ | ++ |
| 2 | Buccal | T1 N0 M0 | 1 | Exophytic | ++ | - | - |
| 3 | Alveolus | T2 N1 M0 | 3 | Growth | ++ | - | - |
| 4 | Floor | T2 N1 M0 | 3 | Growth | ++ | ++ | ++ |
| 9 | Lateral tongue | T2 N0 M0 | 2 | Growth | - | ++ | ++ |
| 5 | Buccal | T4 N1 M0 | 4 | Growth | ++ | - | - |
| 7 | Alveolus | T2 N1 M0 | 3 | Growth | ++ | ++ | ++ |
| 8 | Buccal | T4 N1 M0 | 4 | Exophytic/ulcer | ++ | ++ | - |
| 6 | Alveolus | T4 N1 M0 | 4 | Growth | ++ | - | - |
| 10 | Buccal | T2 N0 M0 | 2 | White/redulcer | - | + | ++ |
| 12 | Alveolus | T4 N0 M0 | 4 | Growth | ++ | - | - |
| 15 | Buccal | T2 N0 M0 | 2 | White/ulcer | ++ | ++ | + |
| 11 | Buccal | T4 N1 M1 | 4 | Growth | ++ | ++ | ++ |

Figure 1. Hierarchical clustering of the 263 genes found as differentially expressed between tumors and normal controls and their possible correlation to the patients clinicopathological data.

*COL4A1*, *HAS3*, *CD47* and *ANGPTL4* (all were upregulated), *S100A1*, *BPIL1* and *WFDC2* (all were downregulated). These genes were selected for verification because four of them (*GJA1*, *MMP1*, *COL4A1* and *S100A1*) were previously reported in oral cancer, while the rest (*HAS3*, *CD47*, *ANGPTL4*, *BPIL1* and *WFDC2*) were not. Aliquots of the same RNA (200-300 ng) used for the microarray hybridization was also used for synthesis of the cDNA, performed with High Capacity cDNA Archive kit (Applied Biosystems, Foster, CA), following the manufacturer's instructions. Real-time PCR was performed with probes for each gene (listed in Table II) using the ABI 7900 HT (Applied Biosystems) and 384 well optical plates (ABI). Each reaction contained 1 $\mu$l cDNA, 5 $\mu$l 2x TaqMan Universal Master mix (Applied Biosystems), 0.5 $\mu$l Taqman AOD probe and $H_2O$ to a final volume of 10 $\mu$l and was run in triplicate. Cycling parameters were 95ºC for 10 min, followed by 40 cycles at 95˚C for 15 sec and 60˚C for 1 min. Serial diluted standards were run on the same plate and the relative standard curve method was used to calculate the gene expression as described elsewhere. ß-actin was used as an endogenous normalization control to adjust for unequal amounts of RNA.

## Results

Among the 15 patients studied [3 (20%) females and 12 (80%) were males], 13 were regular betel quid chewers. Six (40%) were both cigarette smokers and alcohol users, while one (7%) patient was a regular smoker without other habits. The two (14%) non-betel quid chewers smoked and drank alcohol regularly. Two of the 15 cancers were verrucous carcinomas while all the others were of squamous cell type.

*Gene expression profile*. Determination of the gene expression profile of the 15 Sri Lankan patients, by hybridizing the cDNA from the pair-wise normal and tumor samples to the 35k human oligo microarray slides, resulted in a set of 263 genes found as differently expressed between tumors and normal controls. Of these, 190 (72%) genes (Table IIIA) were found as upregulated and 73 (28%) (Table IIIB) were found as downregulated.

Since there are many genes reported as differentially expressed in OSCCs, we searched the Cancer Genome Anatomy Project (CGAP, http://cgap.nci.nih.gov/) for genes related to head and neck cancer, by using the Gene Library

A. Genes upregulated in tumor versus normal controls.

| Gene name/function | Gene symbol | Oligo_ID | Average fold change |
|---|---|---|---|
| Adhesion associated genes | | | |
| Cytokeratin 6A | KRT6A | H300014335 | 19.452 |
| Keratin 17 | KRT17 | H200000678 | 14.011 |
| Collagen I, alpha 1 | COL1A1 | H300015580 | 8.804 |
| Periostin, osteoblast specific factor | POSTN | H300000923 | 8.632 |
| Collagen IV, alpha 1 | COL4A1 | H200012184 | 7.971 |
| Chloride channel | CLCA2 | H200015766 | 6.891 |
| Odd Oz/ten-m homolog 2 | ODZ2 | H200008152 | 6.394 |
| Hyaluronan synthase 3 | HAS3 | H200007172 | 6.024 |
| Keratin 5 | KRT5 | H300019235 | 5.745 |
| Myosin IB | MYO1B | H200012339 | 5.382 |
| Ladinin 1 | LAD1 | H200002708 | 4.909 |
| Serine proteinase inhibitor | SERPINH1 | H200001857 | 4.709 |
| Thy-1 cell surface antigen | THY1 | H200012556 | 4.017 |
| Molecule interacting with Rab13 | MIRAB13 | H300008711 | 3.977 |
| Bicaudal D homolog 2 (Drosophila) | BICD2 | H200019300 | 3.578 |
| CD47 antigen | CD47 | H200006967 | 3.317 |
| Epithelial protein lost in neoplasm beta | EPLIN | H200001929 | 3.263 |
| Development and differentiation enhancing factor 1 | DDEF1 | H200001924 | 3.076 |
| Osteonectin | SPARC | H200011770 | 3.027 |
| Actinin, alpha 1 | ACTN1 | H200012175 | 2.915 |
| Dipeptidyl-peptidase 3 | DPP3 | H300022630 | 2.702 |
| Hypothetical protein similar to actin | ARPC5L | H200018870 | 2.561 |
| CD 44 antigen | CD44 | H300019814 | 2.542 |
| Eukaryotic translation initiation factor 4A | EIF4A1 | H200012822 | 2.399 |
| Plectin 1 | PLEC1 | H200006713 | 2.302 |
| Integrin, beta 1 | ITGB1 | H300017353 | 2.207 |
| Actin, gamma 1 | ACTG1 | H300006234 | 2.138 |
| Tropomyosin 4 | TPM4 | H300013721 | 2.088 |
| Angiogenesis | | | |
| Matrix metalloproteinase 1 | MMP1 | H200007011 | 20.586 |
| Matrix metalloproteinase 3 | MMP3 | H200007019 | 7.764 |
| Plasminogen activator, urokinase | PLAU | H200006377 | 7.280 |
| Endothelial cell growth factor 1 | ECGF1 | H200005893 | 7.054 |
| Hexabrachion | TNC | H200009494 | 6.792 |
| Interleukin 8 | IL8 | H200000156 | 5.645 |
| Angiopoietin-like 4 | ANGPTL4 | H300021145 | 4.930 |
| Jagged 1 | JAG1 | H300002958 | 4.773 |
| Tryptophanyl-tRNA synthetase | WARS | H300012785 | 4.845 |
| Hypoxia-inducible factor 1, alpha subunit | HIF1A | H200014761 | 2.620 |
| Apoptosis | | | |
| Caspase 1 | CASP1 | H200000628 | 7.589 |
| Cofilin 1 (non-muscle) | CFL1 | H200008422 | 2.388 |
| Cell-cell signaling | | | |
| Interferon, alpha-inducible protein (clone IFI-15K) | G1P2 | H300006942 | 18.761 |
| Interferon, alpha-inducible protein 27 | IFI27 | H200017325 | 8.509 |
| Connexin 43 | GJA1 | H200005947 | 3.940 |
| Cerebral cavernous malformation 2 | CCM2 | H200001861 | 2.626 |
| Protein arginine methyltransferase 1 | PRMT1 | H300019121 | 2.458 |
| Cell growth regulation | | | |
| Solute carrier family 16 | SLC16A1 | H200006037 | 7.932 |
| Bone marrow stromal cell antigen 2 | BST2 | H200012083 | 7.175 |
| Serpin peptidase inhibitor | SERPINB5 | H300020178 | 6.169 |

Table IIIA. Continued.

| Gene name/function | Gene symbol | Oligo_ID | Average fold change |
|---|---|---|---|
| Caveolin 2 | CAV2 | H200013312 | 4.587 |
| N-myc (and STAT) interactor | NMI | H300001976 | 4.525 |
| Guanylate binding protein 2 | GBP2 | H200007999 | 4.412 |
| Retinol binding protein 1 | RBP1 | H300001638 | 4.261 |
| Basonuclin 1 | BNC1 | H300005994 | 4.252 |
| Breast cancer metastasis suppressor 1 | BRMS1 | H300002128 | 3.891 |
| RAB31, member RAS oncogene family | RAB31 | H200015253 | 3.864 |
| EBNA1 binding protein 2 | EBNA1BP2 | H300005177 | 3.554 |
| Caveolin 1 | CAV1 | H200005908 | 3.459 |
| Polo-like kinase 2 | PLK2 | H300012787 | 3.348 |
| Karyopherin alpha 2 | KPNA2 | H300007884 | 3.317 |
| Stratifin | SFN | H300006188 | 3.275 |
| Ribosomal protein S6 kinase | RPS6KB2 | H300007163 | 2.963 |
| Non-metastatic cells 1 | NME1 | H200012135 | 2.963 |
| Proliferation-associated 2G4 | PA2G4 | H300010406 | 2.875 |
| Proteasome (prosome, macropain) 26S subunit | PSMD1 | H200000856 | 2.782 |
| Exostoses (multiple) 2 | EXT2 | H200006075 | 2.593 |
| Suppressor of G2 allele of SKP1 | SUGT1 | H200001007 | 2.354 |
| RNA binding motif | RBMS1 | H300004002 | 2.255 |
| Nitric oxide synthase interacting protein | NOSIP | H300022154 | 2.188 |
| Immune response | | | |
| Ig gamma-4 chain C region | IGHG4 | H300000934 | 10.716 |
| Guanylate binding protein 1 | GBP1 | H200005495 | 9.735 |
| 2'-5'-oligoadenylate synthetase 2 | OAS2 | H300002591 | 8.434 |
| Transporter 1, ATP-binding cassette | TAP1 | H300021278 | 5.455 |
| Cathepsin C | CTSC | H300018822 | 4.684 |
| Interferon, alpha-inducible protein | G1P3 | H200016555 | 4.436 |
| Proteasome activator subunit 2 | PSME2 | H300007599 | 3.071 |
| Apolipoprotein L, 3 | APOL3 | H200015760 | 3.048 |
| Interferon induced transmembrane protein 2 | IFITM2 | H300004907 | 2.941 |
| Tumor necrosis factor, alpha-induced protein 1 | TNFAIP1 | H200006250 | 2.933 |
| Integral to membrane | | | |
| Leucine-rich repeat-containing protein 8A | LRRC8 | H200008145 | 3.525 |
| Intracellular signaling | | | |
| Rag C protein | RRAGC | H200011726 | 3.678 |
| Ras homolog gene family, member D | RHOD | H200002449 | 3.101 |
| Diacylglycerol kinase | DGKZ | H200017225 | 2.975 |
| Cornichon homolog 4 | CNIH4 | H300011960 | 2.943 |
| Intracellular transport | | | |
| Tubulin beta-5 | TUBB6 | H200017078 | 3.826 |
| Nuclear transport factor 2 | NUTF2 | H300010247 | 2.337 |
| Metabolism | | | |
| Cathepsin L2 | CTSL2 | H200010239 | 5.169 |
| Similar to glucosamine-6-sulfatases | SULF2 | H200004653 | 5.082 |
| GM2 ganglioside activator protein | GM2A | H200009479 | 4.964 |
| Selenoprotein X, 1 | SEPX1 | H200017458 | 3.870 |
| Carbohydrate (N-acetylglucosamine-6-O) | CHST2 | H200001687 | 3.864 |

A. Continued.

| Gene name/function | Gene symbol | Oligo_ID | Average fold change |
|---|---|---|---|
| Heat shock 105 kDa/110 kDa protein 1 | HSPH1 | H300019008 | 3.845 |
| Prion protein (p27-30) | PRNP | H200005977 | 3.688 |
| Thioredoxin-like 5 | TXNL5 | H200005936 | 3.497 |
| Homo sapiens cDNA FLJ30135 fis, clone BRACE2000061 | BLOC1S2 | H200004282 | 3.230 |
| Small nuclear ribonucleoprotein polypeptide G | SNRPG | H300006350 | 3.104 |
| Adenosine kinase | ADK | H200010643 | 3.042 |
| Mitochondrial ribosomal protein L15 | MRPL15 | H200002720 | 2.947 |
| Heat shock 70 kDa protein 4 | HSPA4 | H300022500 | 2.895 |
| Mitochondrial ribosomal protein L17 | MRPL17 | H200001870 | 2.811 |
| Small nuclear ribonucleoprotein polypeptide A' | SNRPA1 | H300000296 | 2.670 |
| Similar to peptidylprolyl isomerase (cyclophilin)-like 1 | PPIL1 | H300000902 | 2.664 |
| Prefoldin subunit 2 | PFDN2 | H200009859 | 2.650 |
| Small nuclear ribonucleoprotein polypeptide B" | SNRPB2 | H200006957 | 2.626 |
| Chaperonin containing TCP1, subunit 5 (epsilon) | CCT5 | H200000420 | 2.578 |
| Mitochondrial ribosomal protein L14 | MRPL14 | H300010300 | 2.545 |
| Similar to peptidylprolyl isomerase (cyclophilin)-like 1 | PPIL1 | H300002507 | 2.499 |
| Peptidylprolyl isomerase A (cyclophilin A) | PPIA | H300005405 | 2.382 |
| LSM1 homolog, U6 small nuclear RNA associated | LSM1 | H200011772 | 2.289 |
| Mitochondrial ribosomal protein S12 | MRPS12 | H200001864 | 2.273 |
| Nucleolar protein family A, member 3 | NOLA3 | H300001534 | 1.955 |
| Proteasome (prosome, macropain) 26S subunit | PSMC1 | H200000913 | 1.858 |
| Oncogene | | | |
| V-Ha-ras Harvey rat sarcoma viral oncogene homolog | HRAS | H300018518 | 3.298 |
| Met proto-oncogene (hepatocyte growth factor receptor) | MET | H300013056 | 3.141 |
| V-jun sarcoma virus 17 oncogene homolog (avian) | JUN | H200006516 | 1.926 |
| Signal transduction | | | |
| Signal transducer and activator of transcription 2, 113 kDa | STAT2 | H300010660 | 4.436 |
| Lymphocyte antigen 6 complex, locus E | LY6E | H200006433 | 3.491 |
| Tyrosine kinase, non-receptor, 2 | TNK2 | H300022337 | 2.292 |
| Transcriptional regulation | | | |
| Superoxide dismutase 2, mitochondrial | SOD2 | H200018923 | 4.335 |
| Interferon regulatory factor 7 | IRF7 | H300021511 | 3.581 |
| BolA-like protein 2 | BOLA2 | H300001390 | 3.158 |
| SUB1 homolog (S, cerevisiae) | SUB1 | H300000824 | 3.127 |
| BTB and CNC homology 1, basic leucine zipper Transcription factor 1 | BACH1 | H200013953 | 2.967 |
| Ubiquitin cycle | | | |
| Ubiquitin-conjugating enzyme E2S | UBE2S | H300007235 | 4.858 |
| Proteasome (prosome, macropain) subunit, alpha type, 1 | PSMA1 | H200006899 | 3.307 |
| Proteasome (prosome, macropain) subunit, alpha type, 4 | PSMA4 | H200016300 | 3.206 |
| Proteasome (prosome, macropain) subunit, beta type, 2 | PSMB2 | H300007846 | 3.037 |
| Ring finger protein 149 | RNF149 | H300001827 | 2.706 |
| Ubiquity-conjugating enzyme E2L 3 | UBE2L3 | H300004594 | 2.410 |
| Ring-box 1 | RBX1 | H200017552 | 2.236 |
| Genes with other functions | | | |
| SRY (sex determining region Y)-box 20 | SOX15 | H200010702 | 6.417 |
| BA13B9,3 (Novel protein similar to KRT8) | KRT8L1 | H300008020 | 4.819 |
| TLH29 protein precursor | FAM14A | H200010658 | 4.498 |
| Hematological and neurological expressed 1 | HN1 | H300005778 | 4.229 |
| Nuclear RNA helicase, DECD variant of DEAD box family | DDX39 | H200018834 | 3.966 |
| HSPC160 protein | ORMDL2 | H200002233 | 2.948 |

Table IIIA. Continued.

| Gene name/function | Gene symbol | Oligo_ID | Average fold change |
|---|---|---|---|
| CDW92 antigen | CDW92 | H200008388 | 3.832 |
| Brix domain containing 2 | BXDC2 | H300020946 | 2.822 |
| Solute carrier family 7, member 5 | SLC7A5 | H200014300 | 2.783 |
| Signal peptidase complex subunit 3 homolog | SPCS3 | H200004559 | 2.667 |
| Protein phosphatase 1, regulatory (inhibitor) subunit 14B | PPP1R14B | H300013966 | 2.471 |
| Coiled-coil-helix-coiled-coil-helix domain containing 1 | CHCHD1 | H300005351 | 2.329 |
| DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21 | DDX21 | H200007834 | 1.996 |
| Genes with unknown function | | | |
| Epithelial stromal interaction 1 (breast) | EPSTI1 | H200019973 | 7.766 |
| Family with sequence similarity 83, member A | FAM83A | H200016387 | 6.537 |
| Hypothetical protein MGC4677 | LOC541471 | H300002112 | 4.692 |
| Melanoma-derived leucine zipper, extra-nuclear factor | MLZE | H200013079 | 4.471 |
| Hepatitis delta antigen-interacting protein A | CCDC85B | H200005619 | 3.048 |
| Family with sequence similarity 49, member B | FAM49B | H300020526 | 2.927 |
| Chromosome 19 open reading frame 22 | C19orf22 | H200001829 | 2.891 |
| COBW domain containing 1 | CBWD1 | H300020976 | 2.690 |
| Hypothetical protein FLJ10350 | RPRC1 | H200008278 | 2.685 |
| KIAA0842 protein | KIAA0842 | H200005962 | 2.680 |
| Hypothetical protein MGC22793 | MGC22793 | H300005475 | 2.312 |
| TPA regulated locus | TPARL | H200015568 | 2.186 |
| Unknown | - | H300016521 | 16.121 |
| Unknown | - | H300000399 | 11.819 |
| Unknown | - | H300016456 | 10.856 |
| Unknown | - | H300022261 | 9.555 |
| Unknown | - | H300009062 | 6.931 |
| Hypothetical gene supported by AK023162 | - | H200001742 | 6.817 |
| Unknown | - | H300014044 | 6.517 |
| Unknown | - | H300008061 | 5.898 |
| Unknown | - | H300006085 | 5.625 |
| Unknown | - | H300000084 | 5.119 |
| Q5T7K4_HUMAN | - | H300002726 | 4.300 |
| Hypothetical protein MGC4677 | - | H300002113 | 4.006 |
| Unknown | - | H200001872 | 3.982 |
| FLJ46072 | - | H300002815 | 3.688 |
| Unknown | - | H300000271 | 3.266 |
| Unknown | - | H300006679 | 2.888 |
| Unknown | - | H300021607 | 2.751 |
| Unknown | - | H300003367 | 2.620 |
| Unknown | - | H300008569 | 2.549 |
| Hypothetical protein DKFZp586I1420 | - | H200011821 | 2.523 |
| Unknown | - | H300001105 | 2.479 |
| Unknown | - | H300000638 | 2.439 |
| Unknown | - | H300012353 | 2.439 |
| Unknown | - | H300000861 | 2.325 |
| Unknown | - | H300009853 | 2.256 |
| Unknown | - | H300001828 | 2.242 |
| Unknown | - | H300022664 | 2.216 |
| Unknown | - | H300003465 | 2.196 |
| 13kDa differentiation-associated protein | - | H200004675 | 2.162 |
| Unknown | - | H300009555 | 2.158 |
| Unknown | - | H300009420 | 2.045 |
| Unknown | - | H300008478 | 1.987 |
| Unknown | - | H200003368 | 1.971 |
| Unknown | - | H300003763 | 1.867 |

zer tool and the following search criteria: Organism: iens; Library Group: CGAP Libraries; Tissue type: Head and neck; Library Preparation: Any; Tissue Histology: Cancer; Library Protocol: Any. We compared the differentially expressed genes found in our study with the search results (2,500 genes involved in head and neck cancer) and found that 66 of the genes [such as *COL4A1*, *COL1A1*, *PLAU*, *MMP1*, *MMP3*, *ITGB1*, *SPARC* (upregulated), *SLPI*, *EEF1A1* and *KRT19* (downregulated), among others] matched results from the CGAP search. These genes have also been reported in previous studies related to gene expression profiles in oral cancer (14). The genes that were neither listed in the CGAP nor previously reported in OSCCs included *CD47*, *UBE2L3*, *RPS6KB2*, *APOL3*, *RRAGC* (upregulated), *AZGP1*, *CLN3* and *XBP1* (downregulated). To obtain information on gene function, we searched the CGAP database using the Gene Finder tool and the Gene Ontology Browser. Genes of known function (192 of the 263 genes) were found to be related to cellular processes like cell adhesion, cell signaling, angiogenesis, metastasis and metabolism.

To determine the significant changes related to the biological systems, a GO DAG analysis was performed, which resulted in the following terms: transcription regulator activity, structural molecule activity, intracellular signaling cascade, cytoskeleton and signal transduction.

*Hierarchical clustering*. Hierarchical clustering of the 15 samples using the 263 genes found, resulted in several subgroups (Fig. 1). Sample number 1 and 11 (verrucous SCC and an advanced SCC), did not cluster with any of the other cases. The remaining cases clustered in several subgroups, such as samples with tumors of different TNM status, BQ chewers and non-BQ chewers, smokers, non-smokers, drinkers and non-drinkers. Interestingly however, tumors of stage three and four showed a tendency to group together, although one of the samples (number nine, a tumor of stage 2) clustered with these tumors, which might indicate a tendency towards developing an aggressive behavior though presented as stage 2. Although this is an interesting observation, analysis of additional cases is necessary. All patients were smokers and/or BQ chewers with varying types of alcohol habits. The two non-BQ chewers grouped differently, although their tumors were of stage 2. One of the samples (non-betel quid user) clustered in a large subgroup, while the other separated from the rest in a subgroup on a higher level. Patient number 2, diagnosed with a stage 1 tumor, was also distinguishable from the other patients and grouped separately.

*KEGG pathway analysis*. The KEGG pathway analysis (performed with the 192 genes with known functions), showed six pathways where a significant number of the genes found were included. The predominant pathways found were cell communication and integrin-mediated cell adhesion (Fig. 2) where 10 and 11 genes were represented, respectively.

*Quantitative real-time RT-PCR*. Gene expression profile was validated by quantitative reverse transcription for nine genes. We used ß-actin for normalization and determination of the results. A good correlation was found between the microarray data and the RT-PCR results (Fig. 3).

## Discussion

In this study, we analysed gene expression profile in 15 cases of OSCCs from Sri Lanka and correlated the findings to the clinicopathological parameters. We identified 263 genes as differentially expressed between tumors and normal controls where 190 (72%) were upregulated and 73 (28%) were downregulated. Of the genes found, 192 (73%) were of known functions, where 66 of these genes (34%) (such as *COL1A1*, *COL4A1*, *MMP1*, *TNC* and *PLAU*) were previously reported in the CGAP database (http://cgap.nci.nih.gov/). The 192 known genes were found to be implicated in cellular processes like cell growth, cell proliferation, cell signaling and angiogenesis that were suggested to play an important role in oral tumorigenesis (9,22,23). We selected nine genes (*COL4A1*, *GJA1*, *MMP1*, *HAS3*, *CD47*, *ANGPTL4*, *S100A1*, *BPIL1* and *WFDC2*) and verified their expression with quantitative real-time PCR. Both up- and downregulated genes were selected, as well as genes that were either previously reported or not reported in OSCCs.

*COL4A1*, found as upregulated in this study, is the main constituent of the extracellular matrix (ECM) and the basement membrane (BM) (24). It has been suggested that increased collagen synthesis is related to the use of betel quid and oral submucous fibrosis - a precursor condition to oral cancer (25). Tsai *et al* (26) studied gene expression profile in oral cancers from betel quid chewers in Taiwan and reported an increase in *COL4A1* expression. Other studies in OSCCs have also reported an increased expression of this gene (14). In Taiwan, BQ is used without tobacco, but in Sri Lanka tobacco is commonly added to BQ (2). Our findings of increased *COL4A1* expression in the biopsies examined from Sri Lanka are in agreement with the findings from Taiwan, suggesting a possible role of BQ in the upregulation of *COL4A1*. These results indicate that *COL4A1* might be considered as a possible biomarker in BQ related lesions, but further confirmatory studies are needed.

*HAS3*, showing increased expression in this study, encodes the enzyme hyaluronan synthase 3, one of the three enzymes involved in synthesis of hyaluronan, an ECM related glycosaminoglycan associated with proliferation and migration (27). Hyaluronan has been found to interact with *CD44* (found as upregulated in our study) and was suggested to be implicated in tumor growth and metastasis. There are no reports of *HAS3* in relation to oral cancer and this is also the case with *CD47*, a gene encoding a cell adhesion protein (28).

We detected an upregulation of *ANGPTL4* (angiopoietin-like 4, previously not reported in oral cancer), a gene induced by hypoxia and grouped under angiogenesis-related genes (pro-angiogenic) (29). *MMP1* is a metalloproteinase involved in the breakdown of the ECM during angiogenesis, invasion and metastasis (30,31). Our study showed upregulation of *MMP1*, which is in agreement with other experiments in oral cancer (14,26,32). *GJA1* (Connexin43, a gap junction constituent involved in intercellular communication) (33), has been suggested to play an important role in the regulation of cell growth and is associated with cancer development (33). Our findings of upregulation of this gene are supported by one other previous study on OSCC (34). *S100A1*, found as downregulated in our study, is a member of the *S100* gene

Table IIIB. Genes downregulated in tumor versus normal controls.

| Gene name/function | Gene symbol | Oligo_ID | Average fold change |
|---|---|---|---|
| Apoptosis | | | |
| Phospholipid scramblase 3 | PLSCR3 | H300022882 | -3.103 |
| Ceroid-lipofuscinosis, neuronal 3, juvenile | CLN3 | H300012300 | -2.188 |
| Adhesion associated genes | | | |
| Statherin | STATH | H200004368 | -101.425 |
| Alpha-2-glycoprotein 1, zinc | AZGP1 | H200000021 | -41.885 |
| Keratin 19 | KRT19 | H200008573 | -15.716 |
| Cysteine-rich secretory protein 3 | CRISP3 | H200005086 | -12.501 |
| Mucin 5, subtype B, tracheobronchial | MUC5B | H300022927 | -9.895 |
| Keratin 7 | KRT7 | H200003337 | -4.630 |
| Cell growth regulation | | | |
| Putative cytokine high in normal-1 | HIN-1 | H200005487 | -17.842 |
| Tetraspan 1 | TSPAN1 | H200004450 | -9.837 |
| N-myc downstream-regulated gene 2 | NDRG2 | H200015807 | -4.738 |
| Immune response | | | |
| Secretory leukocyte protease inhibitor | SLPI | H200016308 | -9.310 |
| Immunoglobulin J polypeptide | IGJ | H300003753 | -4.258 |
| X-box binding protein 1 | XBP1 | H200013682 | -3.134 |
| Intracellular signaling | | | |
| S100 calcium binding protein A1 | S100A1 | H200009536 | -3.485 |
| Metabolism | | | |
| Phosphoglycerate dehydrogenase | PHGDH | H300006924 | -5.323 |
| Carbonyl reductase | DCXR | H200001845 | -2.262 |
| Eukaryotic elongation factor 1 alpha 1 | EEF1A1 | H300022640 | -2.075 |
| Eukaryotic translation elongation factor 1 gamma | EEF1G | H300009473 | -1.997 |
| Ribosomal protein L22 | RPL22 | H300009237 | -1.929 |
| Similar to ribosomal protein L13a | LOC387841 | H300002118 | -1.672 |
| Ribosomal protein L13 | RPL13 | H300021124 | -1.565 |
| Proteolysis | | | |
| WAP four-disulfide core domain 2 | WFDC2 | H300022511 | -27.078 |
| Kallikrein 11 | KLK11 | H200005275 | -4.398 |
| ATG4 autophagy related 4 homolog B | APG4B | H300003574 | -2.167 |
| Signal transduction | | | |
| Calcitonin-related polypeptide, beta | CALCB | H200017136 | -15.155 |
| Protein FAM3D precursor | FAM3D | H200005421 | -10.170 |
| Tetraspanin 13 | TSPAN13 | H300011863 | -5.236 |
| GNAS complex locus | GNAS | H300022096 | -2.367 |
| Transcriptional regulation | | | |
| Zinc finger protein 253 | ZNF253 | H300008475 | -2.138 |
| Genes with other functions | | | |
| Hypothetical protein dJ726C3.2 | BPIL1 | H200016425 | -196.506 |
| Proline rich 4 (lacrimal) | PRR4 | H300001196 | -41.384 |
| Mucin 7, salivary | MUC7 | H200011272 | -37.298 |
| Trefoil factor 3 (intestinal) | TFF3 | H300004783 | -20.952 |
| Transcobalamin | TCN1 | H200000515 | -9.862 |
| Olfactory receptor 5T2 | OR5T2 | H300003770 | -8.712 |
| Hemoglobin, alpha 1 | HBA1 | H200016940 | -8.655 |
| Crumbs homolog 3 (Drosophila) | CRB3 | H200013695 | -8.483 |
| Hemoglobin, beta | HBB | H300022339 | -5.687 |
| Defensin, beta 1 | DEFB1 | H200004191 | -4.573 |

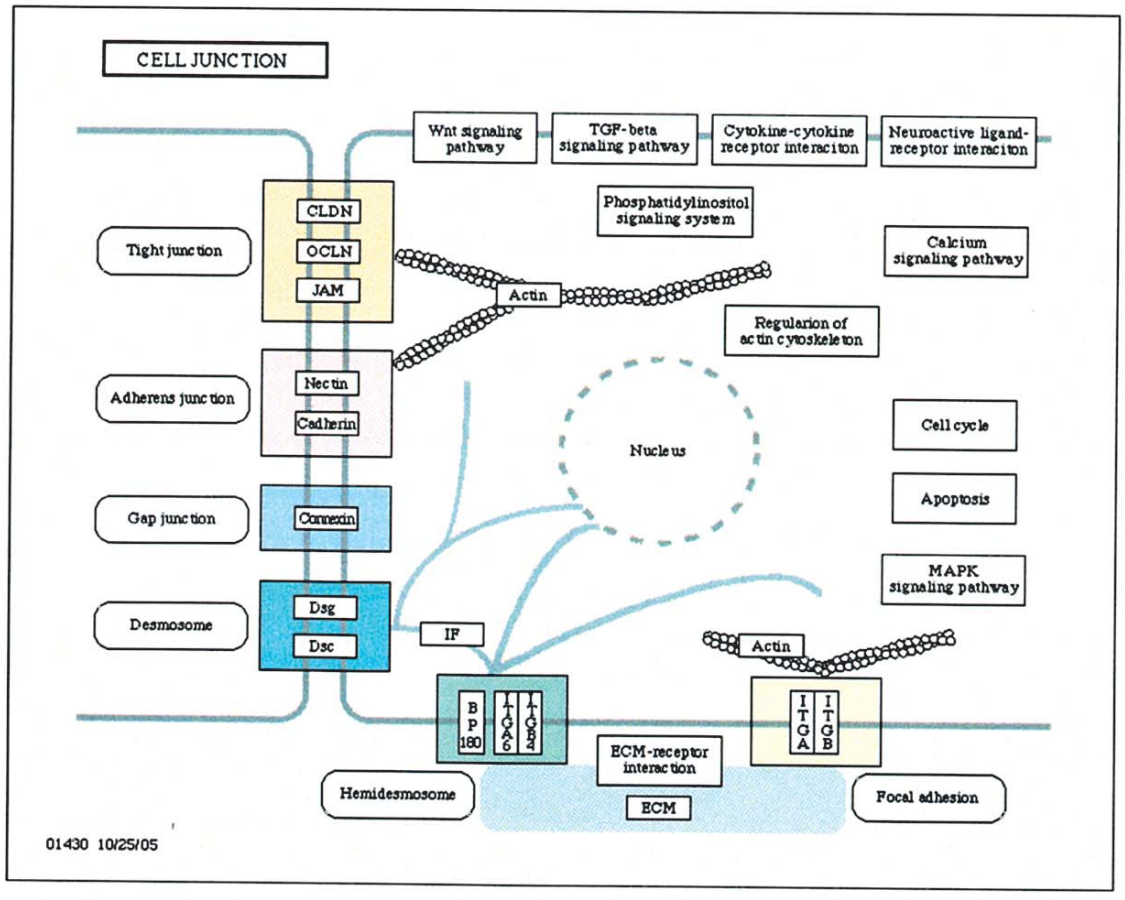
Table 3. Continued.

| Gene name/function | Gene symbol | Oligo_ID | Average fold change |
|---|---|---|---|
| Hypothetical protein FLJ40504 | FLJ40504 | H300004165 | -3.751 |
| Olfactory receptor 8G5 | OR8G5 | H300006949 | -3.561 |
| H2A histone family, member L | HIST1H2AC | H200003862 | -3.446 |
| Lysozyme (renal amyloidosis) | LYZ | H200015514 | -3.077 |
| Transmembrane emp24 protein transport domain containing 3 | TMED3 | H200008352 | -2.719 |
| Reticulon 3 | RTN3 | H300008519 | -2.583 |
| Chromosome 1 open reading frame 8 | C1orf8 | H200002003 | -2.359 |
| ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 2 | ATP5G2 | H200010315 | -2.216 |
| Genes with unknown function | | | |
| Hypothetical protein LOC124220 | LOC124220 | H200011403 | -180.985 |
| Anterior gradient 2 homolog (Xenepus laevis) | AGR2 | H200010467 | -37.251 |
| Dpy-like 2 protein | DPY19L2 | H300011031 | -27.669 |
| Chromosome 4 open reading frame 7 | C4orf7 | H300008107 | -10.921 |
| La ribonucleoprotein domain family, member 2 | LARP2 | H300022914 | -4.066 |
| KIAA1693 | NBPF1 | H300006550 | -3.177 |
| Glioma tumor suppressor candidate region gene 2 | GLTSCR2 | H300001487 | -2.023 |
| Unknown | - | H300007292 | -98.978 |
| Unknown | - | H300015187 | -24.882 |
| MGC34829 protein | - | H300004686 | -23.621 |
| Hypothetical protein DKFZp586L111 | - | H200019198 | -22.175 |
| Unknown | - | H300011211 | -18.810 |
| Hypothetical protein LOC90113 | - | H200019881 | -13.858 |
| PRO1848 | - | H200007208 | -6.556 |
| Hypothetical protein DKFZp781I24108 | - | H300001302 | -4.690 |
| Unknown | - | H300008514 | -4.604 |
| Unknown | - | H300005667 | -4.363 |
| Unknown | - | H300008538 | -4.245 |
| Unknown | - | H300010322 | -3.754 |
| Unknown | - | H300017074 | -2.986 |
| Homo sapiens, clone IMAGE:3633225, mRNA | - | H200020471 | -1.892 |
| Unknown | - | H300009486 | -1.839 |
| Unknown | - | H300004908 | -1.783 |
| Unknown | - | H300002725 | -1.714 |
| Unknown | - | H300000309 | -1.644 |

family, coding for $Ca^{2+}$-binding proteins, suggested to play an important role in cancer progression (35). Downregulation of *S100A1* has been reported in another study on head and neck cancer (36). *WFDC2*, Wap four-disulfide core domain 2 (found as downregulated in this study), is a gene encoding a protease inhibitor and has been suggested as a tumor biomarker in ovarian carcinoma (37).

OSCC is a major health problem related to betel quid use and smokeless tobacco habits in developing countries, albeit there are few studies focusing on gene expression profile of oral cancers from these countries. There are several studies performed in OSCCs from western countries (14,32,38-40), and some of the genes found in our study to be differentially expressed correlate with findings from western countries. These include in particular *COL4A1*, *COL1A1*, *MMP1*, *PLAU*, *SPARC*, *TNC* and *KRT19*, with *MMP1* and *PLAU* as the most frequent genes. MMPs have been described as possible biomarkers of invasion and metastasis in oral cancers (30), which might also apply for the cases examined herein. *PLAU* has been suggested to be implicated in enhanced cell proliferation and migration (41) and as a prognostic marker for relapse-free survival of OSCCs, together with its receptor *u*PAR (42). *SPARC*, or osteonectin, is also implicated in ECM breakdown (43) and has been reported to be associated with tumor progression and metastasis (44). This protein was previously reported in two studies from Sudan, where the use of oral snuff (*toombak*) is common (45,46). The frequent reports of the ECM related genes in association with OSCCs might suggest that ECM breakdown is an important process in oral carcinogenesis. This is supported by our finding of cell adhesion as one of two predominant biological pathways. *MMP1* is related to *COL4A1*, which is an important ECM constituent, suggesting that an increase in *COL4A1* expression might be related to betel quid usage as suggested by others

A




B

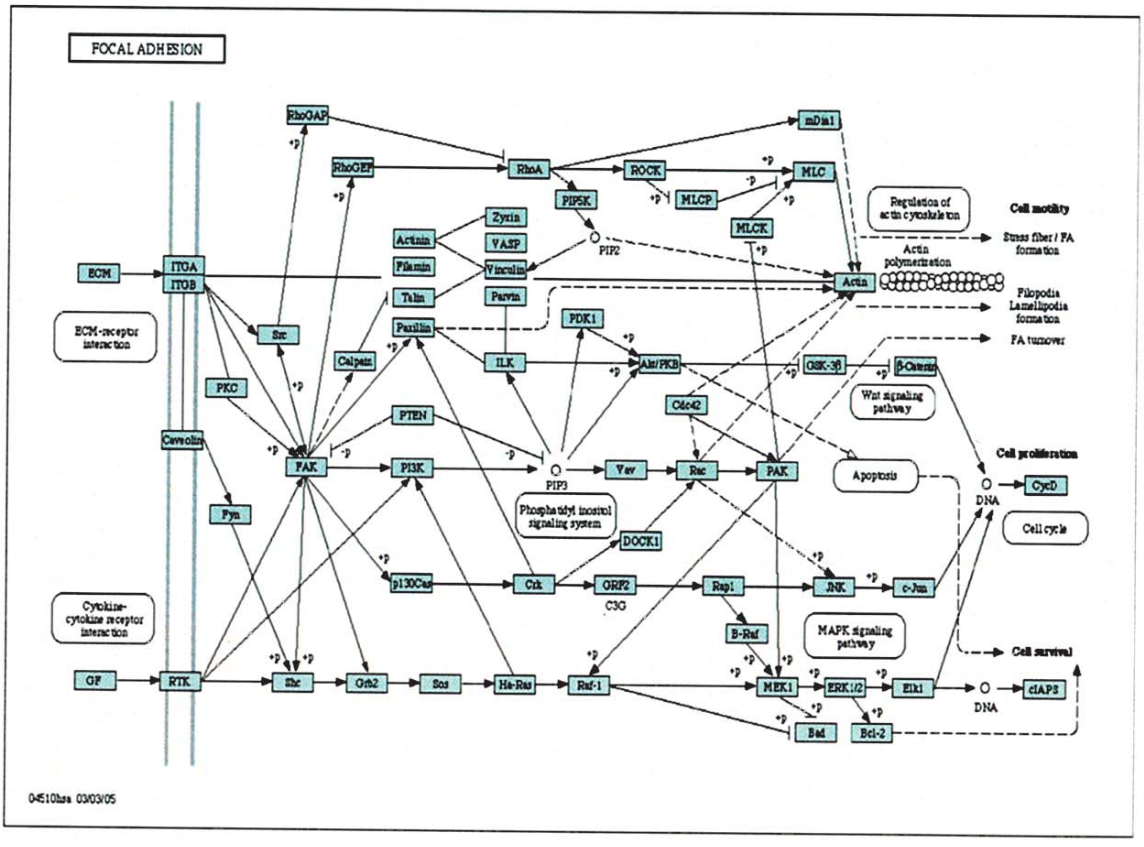



Figure 2. KEGG analysis illustrating cell communication (2A) and focal adhesion molecules (2B) pathways.

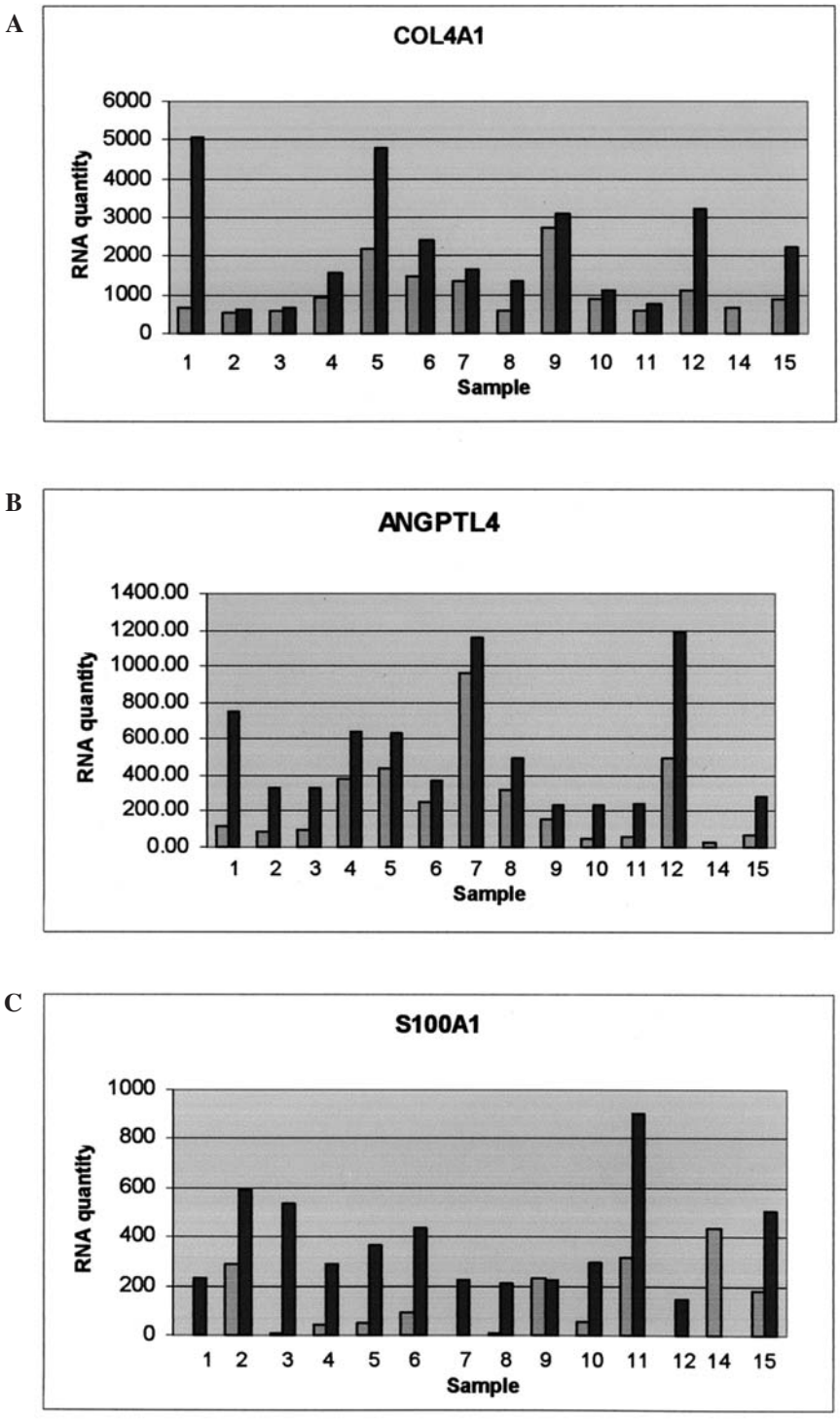



Figure 3. Association between the microarray hybridization and the QRT-PCR data. Results are presented here for COL4A1 (A), ANGPTL4 (B) and S100A1 (C). The light columns, RT-PCR data; the dark columns, microarray data.

(25,47,48). Since *PLAU* may also be related to *COL4A1* through ECM breakdown, *MMP1*, *COL4A1*, *PLAU* and *SPARC* might be suggested as possible biomarkers for OSCCs associated with betel quid chewing as a causative factor. The fact that these four genes have been consistently reported in relation to cancers of the oral cavity in Asian populations might indicate a possible relationship between their increased expression and betel quid usage.

Gene expression profiling not only offers the possibility of finding specific genes differentially expressed between cancer and normal tissue, but individual samples or subgroups can also be distinguished by determining the general profile of each sample. In our study, hierarchical clustering of the analyzed samples showed several subgroups, with the exception of one case diagnosed with a verrucous SCC - a tumor related to smokeless tobacco that is usually less

aggressive compared to other OSCCs (39,40,49). Another sample also differed from the other cases by grouping separately, and this was the only stage 1 tumor among the 15 cancers analyzed. We observed a large subgroup of samples with tumors predominately at stage 3 and 4, which might suggest that advanced tumors have a common gene expression profile compared with tumors at stage 1 and 2. Interestingly however, one other stage 2 tumor clustered with tumors of stage 3 and 4, suggesting a gene expression profile in this sample that is similar to advanced tumors. Relevance of these observations needs to be verified in larger samples. Other OSCC studies using gene expression profiles have elucidated a way of distinguishing metastasizing tumors from non-metastasizing tumors (14). Data of this kind may be valuable in predicting the outcome and risk of metastasis of different tumors, in addition to the traditional tumor classification criteria. Identification of aggressive tumors also offers the possibility of describing individual therapeutic approaches and improved treatment planning.

Further studies will therefore be necessary to confirm our findings, which might lead to a better understanding of the molecular basis of cancer development and tumor progression, particularly in relation to BQ induced oral cancers.

## Acknowledgements

The skilled technical assistance of Ms. Inger Ottesen and Gudveig Fjell is highly appreciated. We thank Bjørn E. Kristiansen, Harald Breilid, Rita Holdhus and Anne-Kristin Stavrum at the Norwegian Microarray Consortium for assistance. Funded by The Norwegian Research Council, Meltzer Høyskolefond, Norsk Dental Depot and UICC ICRETT Fellowship (SOI).

## References

1. Ferlay J, Bray F, Pisani P and Parkin DM: Globocan 2002. Cancer incidence, mortality and prevalence worldwide. IARC Press, 2004.
2. Daftary DK, Murti PR, Bhonsle RB, Gupta PC, Mehta FS and Pindborg JJ: Oral squamous cell carcinoma. In: S Prabhu, D Wilson, D Daftary and N Johnson (eds.), Oral diseases of the tropics. Oxford, University Press, Oxford, pp 429-428, 1992.
3. Chiba I: Prevention of betel quid chewers' oral cancer in the Asian-Pacific area. Asian Pac J Cancer Prev 2: 263-269, 2001.
4. IARC: Betel-quid and areca-nut chewing and some areca-nut-derived nitrosamines., Vol. 85. IARC, Lyon, 2004.
5. Nair U, Bartsch H and Nair J: Alert for an epidemic of oral cancer due to the use of betel quid substitutes *guthka* and *pan masala*: a review of agents and causative mechanisms. Mutagenesis 19: 251-262, 2004.
6. Hecht SS: Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev 3: 733-744, 2003.
7. CancerStats. Oral Cancer UK. In: CRC (ed.). London: Cancer Research UK, 2005.
8. Warnakulasuriya S and Ralhan R: Clinical pathological, cellular and molecular lesions caused by smokeless tobacco. J Oral Pathol Med (In press).
9. Scully C, Field JK and Tanzawa H: Genetic aberrations in oral or head and neck squamous cell carcinoma (SCCHN): 1. Carcinogen metabolism, DNA repair and cell cycle control. Oral Oncol 36: 256-263, 2000.
10. Reibel J: Prognosis of oral pre-malignant lesions: significance of clinical, histopathological and molecular biological characteristics. Crit Rev Oral Biol Med 14: 47-62, 2003.
11. Nagpal JK and Das BR: Oral cancer: Reviewing the present understanding of its molecular mechanism and exploring the future directions for its effective management. Oral Oncol 39: 213-221, 2003.
12. Otero-Rey E, Garcia-Garcia A, Barros-Angueira F, Torres-Español M, Gàndara-Rey JM and Somoza-Martin M: DNA microarrays in oral cancer. Med Oral 9: 288-292, 2004.
13. Russo G, Zegar C and Giordiano A: Advantages and limitations of microarray technology in human cancer. Oncogene 22: 6497-6507, 2003.
14. Choi P and Chen C: Genetic expression profiles and biological pathway alterations in head and neck squamous cell carcinoma. Cancer 104: 1113-1128, 2005.
15. Harris CC: Interindividual variation in binding of benzo[a]pyrene to DNA in cultured human bronchi. Science 194: 1067-1069, 1976.
16. Cawson R and Eveson J: Oral pathology and diagnosis. Gower, London, 1987.
17. Dysvik B and Jonassen I: J-Express: exploring gene expression data using Java. Bioinformatics 17: 369-370, 2001.
18. Efron B, Tibshirani R, Goss V and Chu G: Microarrays and their use in a comparative experiment. Technical Report. Stanford University, Division of Biostatistics, 2000.
19. Kaneisha M: A database for post-genome analysis. Trends Genet 13: 375-376, 1997.
20. Kaneisha M and Goto S: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30, 2000.
21. Kaneisha M, Goto S, Hattori M, Aoki-Kinoshita K, Itoh M, Kawashima S, Katayama T, Araki M and Hirakawa M: From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34: 354-357, 2006.
22. Scully C, Field JK and Tanzawa H: Genetic aberrations in oral or head and neck squamous cell carcinoma 2: chromosomal aberrations. Oral Oncol 36: 311-327, 2000.
23. Scully C, Field JK and Tanzawa H: Genetic aberrations in oral or head and neck squamous cell carcinoma 3: clinico-pathological applications. Oral Oncol 36: 404-413, 2000.
24. Stevens A and Lowe J: Human Histology. Elsevier Ltd., 2005.
25. Rajalalitha P and Vali S: Molecular pathogenesis of oral submucous fibrosis - a collagen metabolic disorder. J Oral Pathol Med 34: 321-328, 2005.
26. Tsai W-C, Tsai S-T, Ko J-Y, Jin Y-T, Li C, Huang W, Young K-C, Lai M-D, Liu H-S and Wu L-W: The mRNA profile of genes in betel quid chewing oral cancer patients. Oral Oncol 40: 418-426, 2004.
27. Menzel E and Farr C: Hyaluronidase and its substrate hyaluronan: biochemistry, biological activities and therapeutic uses. Cancer Lett 131: 3-11, 1998.
28. Brown E and Frazier W: Integrin-associated protein (CD47) and its ligands. Trends Cell Biol 11: 130-135, 2001.
29. Le Jan S, Amy C, Cazes A, Monnot C, Lamandè N, Favier J, Philippe J, Sibony M, Gasc J, Corvol P and Germain S: Angiopoietin-like 4 is a proangiogenic factor produced during ischemia and in conventional renal cell carcinoma. Am J Pathol 162: 1521-1528, 2003.
30. Franchi A, Santucci M, Sardi I, Paglierani M and Gallo O: Expression of matrix metalloproteinase 1, matrix metalloproteinase 2 and matrix metalloproteinase 9 in carcinoma of the head and neck. Cancer 95: 1902-1910, 2002.
31. Thomas GT, Lewis MP and Speight PM: Matrix metalloproteinases and oral cancer. Oral Oncol 35: 227-233, 1999.
32. Toruner G, Ulger C, Alkan M, Galante AT, Rinaggio J, Wilk R, Tian B, Soteropoulos P, Hameed MR, Schwalb MN and Dermody JJ: Association between gene expression profile and tumor invasion in oral squamous cell carcinoma. Cancer Genet Cytogenet 154: 27-35, 2004.
33. Mesnil M: Connexins and cancer. Biol Cell 94: 493-500, 2002.
34. Jeon GA, Lee JS, Patel V, Gutkind J-S, Thorgeirsson SS, Kim EC, Chu I-S, Amornphimoltam P and Park MH: Global gene expression profiles of human head and neck squamous carcinoma cell lines. Int J Cancer 112: 249-258, 2004.
35. Emberley E, Murphy L and Watson P: S100 proteins and their influence on pro-survival pathways in cancer. Biochem Cell Biol 82: 508-515, 2004.
36. Jeong W, Chang T-S, Boja ES, Fales HM and Rhee SG: Roles of TRP14, a thioredoxin-related protein in tumor necrosis factor-{alpha} signaling pathways. J Biol Chem 279: 3151-3159, 2004.
37. Hellström I, Raycraft J, Hayden-Ledbetter M, Ledbetter J, Schummer M, McIntosh M, Drescher C, Urban N and Hellström K: The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. Cancer Res 63: 3695-3700, 2003.
38. Arora S, Matta A, Shukla NK, Deo SVS and Ralhan R: Identification of differentially expressed genes in oral squamous cell carcinoma. Mol Carcinog 42: 97-108, 2005.

Iwaran N, Wu J, Sacks P, Gilcrease M and Zacharias W: Array gene expression profiling of cell lines from primary and metastatic tongue squamous cell carcinoma: possible insights from emerging technology. J Oral Pathol Med: 34: 77-86, 2005.
40. Gottschlich S, Ambrosch P, Cordes C, Görögh T, Schreiber S and Häsler R: Gene expression profiling of head and neck squamous cell carcinoma using cDNA microarrays. Int J Oncol 29: 605-613, 2006.
41. Duffy MJ and Duggan C: The urokinase plasminogen activator system: a rich source of tumor markers for the individualised management of patients with cancer. Clin Biochem 37: 541-548, 2004.
42. Hundsdorfer B, Zeilhofer H-F, Bock KP, Dettmar P, Schmitt M, Kolk A, Pautke C and Horch H-H: Tumor-associated urokinase-type plasminogen activator (uPA) and its inhibitor PAI-1 in normal and neoplastic tissues of patients with squamous cell cancer of the oral cavity - clinical relevance and prognostic value. J Craniomaxillofac Surg 33: 191-196, 2005.
43. Chin D, Boyle GM, Williams RM, Ferguson K, Pandeya N, Pedley J, Campbell CM, Theile DR, Parsons PG and Coman WB: Novel markers for poor prognosis in head and neck cancer. Int J Cancer 113: 787-797, 2004.
44. Kato Y, Nagashima Y, Baba Y, Kawano T, Furukawa M, Kubota A, Yanoma S, Imagawa-Ishiguro Y, Satake K, Taguchi T, Hata R, Mochimatsu I, Aoki I, Kameda Y, Inayama Y and Tsukuda M: Expression of SPARC in tongue carcinoma of stage II is associated with poor prognosis: An immunohistochemical study of 86 cases. Int J Mol Med 16: 263-268, 2005.
45. Dysvik B, Vasstrand EN, Løvlie R, Elgindi OAA, Kross KW, Aarstad JH, Johannessen AC, Jonassen I and Ibrahim SO: Gene expression profiles of head and neck carcinomas from Sudanese and Norwegian patients reveal common biological pathways regardless of race and lifestyle. Clin Cancer Res 12: 1109-1120, 2006.
46. Ibrahim SO, Aarsaether N, Holsve MK, Kross KW, Heimdal JH, Aarstad JH, Liavaag PG, Elgindi OAA, Johannessen AC, Lillehaug JR and Vasstrand EN: Gene expression profile in oral squamous cell carcinomas and matching normal oral mucosal tissues from black Africans and white Caucasians: the case of the Sudan vs. Norway. Oral Oncol 39: 37-48, 2003.
47. Chen J, He Q, Yuen A and Chiu J: Proteomics of buccal squamous cell carcinoma: The involvement of multiple pathways in tumorigenesis. Proteomics 4: 2465-2475, 2004.
48. Tilakaratne WM, Klinikowski MF, Saku T, Peters TJ and Warnakulasuriya KAAS: Oral submucous fibrosis: Review on aetiology and pathogenesis. Oral Oncol 42: 561-568, 2006.
49. Warnakulasuriya KAAS and Ralhan R: Clinical, pathological, cellular and molecular lesions caused by oral smokeless tobacco-a review. J Oral Pathol Med 36: 63-77, 2007.