

# **Manuscript 5**

## **Gene Expression Profile of Oral Squamous Cell Carcinomas from Sri Lankan Betel Quid Users**

Mai Lill Suhr<sup>1</sup>, Bjarte Dysvik<sup>2</sup>, Ove Bruland<sup>3</sup>, Saman Warnakulasauriya<sup>4</sup>, Asoka N. Amaratunga<sup>5</sup>,  
Inge Jonassen<sup>2,6</sup>, Endre N. Vasstrand<sup>7</sup>, and Salah O. Ibrahim<sup>1\*</sup>

<sup>1</sup>*Department of Biomedicine, University of Bergen, Jonas Lies vei 91, N5009 Bergen, Norway*

<sup>2</sup>*Department of Informatics, University of Bergen, HIB, N5020 Bergen, Norway*

<sup>3</sup>*Center of Medical Genetics and Molecular Medicine, Haukeland University Hospital, N5021 Bergen, Norway*

<sup>4</sup>*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*

<sup>5</sup>*Faculty of Dental Sciences, University of Peradeniya, Peradeniya 20400, Sri Lanka*

<sup>6</sup>*Computational Biology Unit, BCCS, HIB, N5020 Bergen, Norway*

<sup>7</sup>*Dental Faculty-Periodontology, University of Bergen, Årstadveien 17, N5009 Bergen, Norway*

**Running title** (Gene expression in oral cancers)

### **Keywords:**

Oral squamous cell carcinoma, Microarrays, Genomics, Betel quid, Sri Lanka

**\*Correspondence**

**Summary:**

The incidence of oral squamous cell carcinoma (OSCC) in Sri Lanka is one of the highest in the world, making this cancer a major health problem in this country. Betel quid chewing, a common oral habit in Sri Lanka, is linked to the observed high incidence of OSCCs. In this work, we examined gene expression profile of 15 patients from Sri Lanka, using OSCC samples and pair-wised normal controls, and correlated the findings with clinicopathological parameters. RNA was extracted, cDNA synthesized and hybridised to 35k human oligo DNA microarrays. Following hybridization, scanning and data analysis, 263 genes (190 (72%) were up-regulated and 73 (28%) were down-regulated) were found to be differentially expressed between cancers and normal controls. 66 (25%) of the genes found were novel genes. 66 (25%), including , *MMP3*, *COL4A1*, *PLAU* and *KRT19*, among others, were previously reported in oral cancer. Interestingly, however, 66 of the genes with known function, including *BRMS1*, *CD47* and *CAV2*, were among the genes not previously reported in OSCC. Signal transduction, intracellular signaling and cell membrane molecules were predominant upon GO analysis of the 197 known genes. Of biological pathways involved, the cell communication and integrin-mediated cell adhesion pathways showed as predominant. Hierarchical clustering of the 263 genes found demonstrated separate grouping of two patients (one with verrucous OSCC and the other with advanced disease) from the rest of the patients that were grouped in mixed subgroups with mixed clinicopathological parameters. Results for 4 genes were validated by quantitative real-time PCR. To conclude, the current study is the first report on gene expression profile for Sri Lanka. Valuable information and suggested candidate known and novel genes obtained. We suggest biological pathways of cell communication and integrin-mediated cell adhesion as pathways involved in oral cancers associated with betel quid chewing.

## Introduction

Oral squamous cell carcinoma (OSCC) is a major health problem in developing countries, with the highest incidence found in countries where the prevalence of the use of smokeless tobacco and/or areca-betel quid is high<sup>1</sup>. Of the 500,000 new cases of OSCCs reported world-wide annually, 62% occur in developing countries with the highest rates reported from Sri Lanka and other parts of Southeast Asia<sup>1</sup>. In this part of the globe an estimated 600 million people are reported to chew areca nut and betel quid (BQ)<sup>2</sup>. In Sri Lanka, OSCC is linked to the oral habit of BQ use<sup>3</sup>. 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<sup>4</sup>. 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<sup>5</sup>. BQ-SNAs includes *N*-nitroso-guvacoline (NG) and 3-(methyl-*N*-nitrosamino)propionitrile (MNPN) that is also found in tobacco<sup>6</sup>. In Sri Lanka, fresh areca nut, slaked lime from seashells, betel leaf and dried (or processed) tobacco are chewed with the quid parked between the cheek mucosa, the lower gingiva, and sometimes retained during sleep<sup>2</sup>. Buccal mucosa represents the primary site for OSCC among betel quid chewers<sup>7</sup>, contrary to tongue cancer, which represent the primary site of cancers in Western countries where cigarette smoking and heavy alcohol consumption are the main causative factors<sup>8</sup>. Molecular mechanisms and steps of carcinogenesis in betel quid induced oral cancers may differ from those caused by smoking tobacco use<sup>9</sup>.

Development of OSCC is a multistep process involving genes related to cell cycle, growth control, apoptosis, DNA damage response and other cellular regulators<sup>10</sup>. Understanding the genetic processes and biological pathways involved in the development of OSCC might lead to better methods of disease classification, early detection and diagnosis, as well as therapeutic planning and drug development<sup>11, 12</sup>. cDNA microarrays represents a promising tool that makes it possible to explore the expression pattern of thousands of genes simultaneously, at the RNA level<sup>13, 14</sup>. In the literature, there are several microarray studies on OSCC with promising findings<sup>15, 16</sup>. Although the influence of life-style factors such as nutritional, alcohol use are important to consider in the causation of OSCCs, there is a wide inter-individual difference(s) in susceptibility to chemical carcinogens<sup>17</sup>. Since BQ is associated with OSCC development, we applied 35k human oligo microarrays to examine 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  $\pm 8,68$ ) with previously untreated OSCCs operated on at the Department of Maxillofacial Surgery, Dental School, University of Peradeniya, Sri Lanka. The study has been approved by the Ethical committee at the University of Peradeniya. Normal control samples were obtained either from the opposite side of the tumor or from an area that was at least 4-5 cm away from the cancerous tissue, and was microscopically normal. After surgery, tissue samples (malignant and normal), were immediately submerged in the tissue storage and RNA stabilization solution, *RNAlater*<sup>™</sup> (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 tumours 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 tumours were histologically graded into high, moderate or poorly differentiated carcinomas according to Cawson and Eveson<sup>18</sup>. To rule out gene expression alterations because of stromal cell contamination, we confirmed pathologically that each tumor specimen contained  $\geq 70\%$  cancer cells by analysis of the corresponding H&E-stained sections. For all the patients, data on clinicopathological parameters, including information on betel quid chewing, were available as shown in Table 1.

### *Tissue samples and RNA extraction*

Total RNA was extracted from both tumor and normal controls using TRIzol<sup>®</sup> reagent (Gibco BRL, Carlsbad, CA, USA)/RNeasy Fibrous Tissue Kit (Qiagen Inc., Valencia, CA, USA) following manufacturer's instructions. Quality and quantity of the RNA were determined spectrophotometrically

with a Beckman DU<sup>®</sup>530 Life Science UV/Vis 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). Synthesized cDNA was labeled with Cy<sup>TM</sup>3 (normal cDNA) and Cy<sup>TM</sup>5 (tumour 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), obtained from Operon (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](http://www.mikromatrise.no)). Labeled cDNA was hybridized on the Ventana Discovery<sup>®</sup> XT System (Ventana Medical Systems Inc., Tucson, AZ, USA) according to manufacturer's protocols. 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 quantitation 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](http://www.molmine.com))<sup>19</sup>. Each array was first pre-processed separately by performing the following steps: (1) Spots flagged by Genepix ("bad", "absent" or "not found", -100, -75 and -50, respectively) were filtered, (2) 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, (3) a global lowess normalization was applied to all values left after the filtering step. Thereafter, all in-array replicated 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 tumour versus 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) have no value were removed. Missing values were set to zero, thus avoiding their contribution in indicating up- or down-regulation status of a gene in tumour versus normal control. To prepare the expression matrix for array comparison, we applied scale normalization to reduce differences in expression spread. For finding a gene with a significant difference between tumour and normal control, the relative difference in gene expression  $S_g = M_g / (s_g + a_0)$  was used, where  $M_g$  is the mean log ratio,  $s_g$  is the standard deviation of the log ratios, and  $a_0$  is an added constant for all genes.  $S_g$  is a student's t-statistics with a fudge factor  $a_0$  which correct for underestimated variances and assign a higher weight to high average fold change versus low variance that can be justified by noisy nature of microarray experiments. In our case, and as suggested by Efron, Tibshirani, Goss and Chu<sup>20</sup>,  $a_0$  is set to a 90 percentile of all gene standard errors  $SE_i$ .

Since all tumours labeled with Cy5, and corresponding normal controls with Cy3, we included an additional set of hybridizations of 5 pairs of experiments that contained the same biological samples hybridized twice with a dye swap 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 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 chose 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 tumours and normal controls or alternatively as a result of 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 s-score above 0.5 in the dye swap expression matrix resulting in 263 genes that we believe are differentially expressed between tumours and normal controls. We further performed a permutation experiment and generated 1000 permuted matrices (containing 12034 unfiltered genes) in each of which we randomly flipped the sign of some

of the columns. On average, the 1000 permuted matrices generated 0.47 genes with an s-score above 1.0 resulting in to 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 DAG and thereafter comparison of the number of the selected genes (263) in GO terms was done to the number of genes in a GO DAG based on all genes left from low-level data preparation (12034 genes). Fisher-Irwin exact test was used to calculate a *p*-value for all GO terms using a *p*-value cutoff at 0.1. All terms with less than 3 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 (Kyoto Encyclopedia of Genes and Genomes, [www.KEGG.com](http://www.KEGG.com)) analysis. We also used Fisher-Irwin exact test to find statistically significant pathways.

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

#### *Quantitative Real-Time RT-PCR*

To validate gene expression profile for selected genes, real-time quantitative RT-PCR was performed for four selected genes: *GJAI*, *MMP1*, *COL4A1* (all up-regulated) and *S100A1* (down-regulated), using aliquots of the same RNA used for the microarray experiments. Three of these genes (*GJAI*, *MMP1* and *COL4A1*) were selected because they were also found as candidate genes in a similar study performed with oral cancers from Sweden and UK (*unpublished data*), while the *S100A1* gene was selected because of its role in cancer-related activity. RNA (200-300 ng) was used for synthesis of the cDNA using High Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA), following the manufacturer's instructions. Real-time PCR was performed with probes for *GJAI* (Hs00748445\_s1), reporter sequence: GACCAGTGGTGCGCTGAGCCCTGC; *MMP1* (Hs00233958\_m1), reporter sequence: TAAAGACAGATTCTACATGCGCACA; *COL4A1* (Hs00266237\_m1) reporter sequence: CCTCCAGGCCTCCCTGTACCTGGGC; *S100A1*(Hs00196704\_m1), reporter sequence:



CCAGGCCAACCGTGCACTGCTGCAA; and endogenous control  $\beta$ -actin (Hs99999903\_m1), reporter sequence: GCCTCGCCTTTGCCGATCCGCCGCC, using the ABI 7900 HT (Applied Biosystems) and 384 well optical plates (ABI). Each individual reaction contained 1  $\mu$ l cDNA, 5  $\mu$ l 2xTaqMan Universal Master mix (Applied Biosystems), 0.5  $\mu$ l Taqman AOD probe and H<sub>2</sub>O to a final volume of 10  $\mu$ l, and was run in triplicate. Cycling parameters were 95°C for 10 min, followed by 40 cycles of 95°C for 15 s 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 relative gene expression, as described<sup>21</sup>.  $\beta$ -actin was used as endogenous normalization control to adjust for unequal amounts of RNA.

## **Results**

Among the 15 patients studied, 3 (20 %) were females and 12 (80 %) were males. With the exception of two cases, all were regular betel quid chewers, 6 (40 %) were both cigarette smokers and alcohol users, while one (7 %) patient was a regular smoker without other habits. Two (13%) of the non-betel quid chewers smoked and drank alcohol regularly. 2 of the 15 cancers were verrucous carcinomas while all others were of squamous type.

### *Gene expression profile*

In this study, gene expression profiles of oral cancers from 15 patients from Sri Lanka were investigated by hybridizing cDNA from pair-wise normal and tumour samples to the 35k human oligo microarray slides. Two-hundred and sixty-three genes were found as differently expressed between tumors and normal controls. Of these, 190 (72%) genes (Table 2a) were up-regulated, and 73 (28%) (Table 2b) were down-regulated.

Since there are many genes reported differentially expressed in head and neck cancer, 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 Summarizer tool and the following search criteria: Organism: Homo sapiens; 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 (2500 genes involved in head and neck cancer), and

66 of our genes (including *COL4A1*, *COL1A1*, *PLAU*, *MMP1*, *MMP3*, *ITGB1*, *SPARC* (all up-regulated), *SLPI*, *EEF1A1* and *KRT19* (all down-regulated), among others) matched the CGAP search results. These genes have also been reported in other gene expression profile studies with oral cancer<sup>22</sup>. Of the genes not listed in CGAP, and not previously reported in OSCC were *BRMS1*, *CD47* and *CAV2*. A Gene Ontology analysis related the 263 genes found to different molecular and tumourigenic processes. Gene function was determined by searching the CGAP database, using the Gene Finder tool and the Gene Ontology Browser. The most predominant functions were related to signal transduction (171 genes), cell membrane (149 genes), integral to membrane (124 genes) and intracellular signaling cascade (105 genes). Of the 263 genes, 66 were of unknown function.

#### *Hierarchical clustering*

Upon hierarchical clustering of the expression profile of the 263 genes, several subgroups were found (Figure 1). Patient number 1, diagnosed with verrucous OSCC, did not cluster within any of the other cases. In the first subgroup, case 11 parted from all other cases. This patient was the only one with both a nodal and distant metastasis. The remaining cases grouped in several subgroups, including patients with tumours of different stage, both betel-quid chewers and non-betel quid chewers, smokers, non-smokers, drinkers and non-drinkers. The tumour grade of these patients varied from 1 to 4, some were diagnosed with nodal metastases. All were smokers or betel quid chewers with different alcohol habits.

#### *KEGG pathway analysis*

The KEGG pathway analysis (performed with 197 genes with known functions), showed six pathways where a significant number of the 263 genes found were included. The predominant pathways found were Cell communication and Integrin-mediated cell adhesion (Figure 2 and 3) where 10 and 11 genes were represented, respectively.

#### *Quantitative Real-Time RT-PCR*

Gene expression was validated by quantitative reverse transcription. We used  $\beta$ -actin for normalization and determination of the results. For comparison to microarray data, the hybridization results for the tumour samples (Cy5) were extracted and normalized using J-Express. Further analysis was done with Microsoft Excel, and the results were evaluated by two of the authors (O.B. and M.L.S). The microarray data and the RT-PCR results correlated well for the expression of the four genes tested (Fig 4).

## Discussion

Application of the microarray technology for mapping of differentially expressed genes between oral cancer and non-cancer cells has identified genes possibly involved in development of this disease<sup>13, 22</sup>. The technology has also been used to explore differences in gene expression between cancers and precancers, and between tumours of different stages<sup>11, 23-25</sup>. In this work, we studied gene expression profile in 15 cases of OSCCs from Sri Lanka, and correlated the findings to clinicopathological parameters. We identified 263 genes as differentially expressed, 190 (72%) as up-regulated and 73 (28%) as down-regulated, between tumours and normal controls. Of the genes found, 197 (75%) (including *MMP1*, *COL1A1*, *COL4A1*, *TNC*, *PLAU*, *KRT 17* and *KRT19*, among others) were of known function, with 66 (25%) (including *MMP1*, *KRT19*, *TNC*, *PLAU* and *CAV2*) being reported earlier in head and neck cancer, as recorded in CGAP database (<http://cgap.nci.nih.gov/>). The 197 genes of known function were implicated in cellular processes like cell growth, cell proliferation, cell signalling and angiogenesis, all suggested to play an important role in oral tumourigenesis<sup>10, 26, 27</sup>. Expression of four genes (*COL4A1*, *GJAI*, *MMP1* and *S100A1*) was validated with quantitative real-time PCR. *COL4A1*, found as up-regulated in this work, is the main constituent of ECM and the basement membrane (BM)<sup>28</sup>. It has been suggested that increased collagen synthesis is related to use of betel quid and oral submucous fibrosis<sup>29</sup>. Tsai et al<sup>30</sup> studied gene expression profile in oral cancers from betel quid chewers in Taiwan, and reported an up-regulation of *COL4A1*. Other microarray studies with oral cancer has also reported up-regulation of this gene<sup>22</sup>. In Taiwan, BQ is used without tobacco, but in Sri Lanka it is used with tobacco. Our findings of up-regulation of *COL4A1* in the

cases examined from Sri Lanka, which are in agreement with findings from Taiwan, suggests a possible role of BQ in up-regulation of *COL4A1* expression. These findings might indicate that *COL4A1* can be considered as a possible candidate biomarker in BQ related lesions. These findings, however, warrant further studies to validate the role of *COL4A1* in BQ related lesions from South East Asia. *GJAI* (Connexin43) is a gap junction constituent involved in intercellular communication<sup>31</sup>. The *GJAI* protein has been suggested to play an important role in regulation of cell growth, and is associated with cancer development<sup>31</sup>. Our findings of up-regulation are supported by another study<sup>32</sup>. *MMP1* is a metalloproteinase involved in breakdown of the extracellular matrix (ECM) during angiogenesis, invasion and metastasis<sup>33, 34</sup>. We found this gene as up-regulated in our study, which is in agreement with other microarray experiments involving oral cancer<sup>25, 30, 35, 36</sup>. *COL4A1*, *GJAI* and *MMP1* were selected not only for their function in carcinogenesis, but also because they were found in another gene expression study that we performed with oral cancers from Sweden and UK (unpublished data).

*S100A1*, found as down-regulated in our study, is a member of the *S100* gene family, coding for Ca<sup>2+</sup>-binding proteins, suspected to play a role in cancer progression<sup>37</sup>. These proteins are implicated both in extracellular and intracellular processes, like immune response, growth and Ca<sup>2+</sup> homeostasis<sup>37</sup>. *S100A1* has been shown to interact with *S100A4*, another gene in the *S100* family, suggested to play a role in metastasis and invasion<sup>37, 38</sup>. Wang et al<sup>38</sup> showed that *S100A1* can act as an antagonist on *S100A4*, reducing its cancer-related activities.

Hierarchical clustering of the 263 genes found resulted in grouping of the cases examined in several subgroups, with the exception of one patient diagnosed with a verrucous OSCC. This is an invasive, but less aggressive cancer than other OSCCs<sup>39</sup>. It is related to use of smokeless tobacco, and has been reported to occur in relation to betel quid chewing<sup>40</sup>. One other patient that did not cluster in any subgroup, was a case with advanced cancer. A relationship between gene expression profiles and metastasis in oral cancer has been suggested by others<sup>25, 23, 41</sup>. There were two patients with no history of betel quid chewing included in the study. One of these patients stood out from the rest in one separate subgroup, while the other grouped with a BQ chewer. One patient, diagnosed with a small tumour, also was distinguishable from other patients, and grouped separately. With the exception of

these cases, there was no obvious relationship between the clustering of the OSCCs examined and the corresponding clinicopathological data. This might suggest that genetic alterations leading to the development of OSCC occur at an early stage during carcinogenesis of this disease. However, and although our clustering results correlates with previous studies on gene expression profiling of oral cancers<sup>30,42</sup>, it is of note that the sample size of our study is too small to draw any certain conclusions.

We performed a KEGG pathway analysis with the 263 differentially expressed genes, which resulted in 6 pathways. The predominant ones were cell communication and integrin-mediated cell adhesion. These pathways are related to cell signalling, cell growth and proliferation, all important for oral tumourigenesis<sup>10, 26, 27, 43</sup>.

Although OSCC is a major health problem related to betel quid use and smokeless tobacco habits in developing countries, there are few studies focusing on gene expression profile of oral cancers from these countries. There are, however, several studies performed in OSCCs from western countries<sup>22, 23, 25, 44, 45</sup>. Some of the genes found in our study as up-or down-regulated correlates with other findings from cases studied from western countries. These include in particular *COL4A1*, *COL1A1*, *MMPI*, *PLAU*, *TNC* and *KRT19*<sup>22, 32, 35, 46, 47</sup>. *MMPI* and *PLAU* were the two most frequent of these genes. MMPs has been described as possible biomarkers of invasion and metastasis in oral cancer<sup>33</sup>, which might also apply for the OSCCs examined herein. The function of *PLAU* is similar. In addition, *PLAU* has been suggested implicated in enhanced cell proliferation and migration<sup>48</sup>. *PLAU* is suggested as a prognostic marker for relapse-free survival of OSCCs, together with its receptor *uPAR*<sup>49</sup>. *MMPI* is related to *COL4A1*, because *MMPI* is involved in ECM breakdown. *COL4A1* is an important ECM constituent, and an increase in *COL4A1* expression might be related to betel quid usage as suggested by our study and others<sup>29, 50, 51</sup>. *PLAU* may also be related to *COL4A1* through ECM breakdown. Thus, *MMPI*, *PLAU* and *COL4A1* might be suggested as possible biomarkers for OSCC development with betel quid chewing as a causative factor. The fact that all three genes have been studied and reported in relation to cancer of the oral cavity, might indicate a possible relationship between their increased expression and betel quid usage.

For conclusion, our results presents here for the first time a gene expression profile study on OSCCs from betel quid chewers from Sri Lanka, and contributes to valuable information related to

grouping of the patients according to clinicopathological data. The differentially expressed genes found included both genes already implicated in oral cancers and other cancers, novel genes, and also genes with unknown function. Genes that might be of interest includes among others *MMP1*, *COL4A1*, *PLAU*, *KRT19*, *TNC*, *S100A1* and *HIN-1*. Further studies on other candidate genes are necessary to understand the role of biological pathways involved in oral cancer tumourigenesis. Our study suggests *MMP1*, *PLAU* and *COL4A1* as possible biomarkers in OSCC related to betel quid chewing.

Sample	Tumour site	TNM (Grade)	Clinical	BQ	Cigarette		Alcohol	
					smoking	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	++	++	++	++	
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 + OSF	++	-	-	-	
15	Buccal	T2 N0 M0 (2)	White/ulcer	++	-	-	+	

Table 1: Patients' clinicopathological data. TNM=Tumour, Node, Metastasis. T.G= Tumour Grade, BQ=Betel quid usage, Regular: ++, occasionally: +, non-user: -

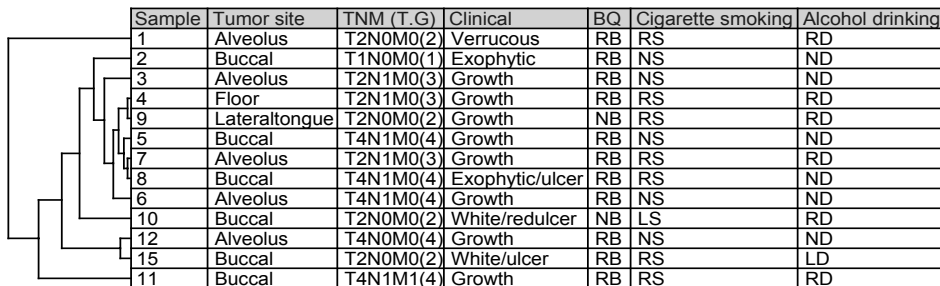


Figure 1: Hierarchical clustering of the gene expression profiles of the 242 genes differentially expressed between tumour tissue and normal tissue in the 13 patients, correlated to the clinicopathological data for all patients. The 13 cases are grouped in several subgroups, as far as two cases (number 1 and 11).

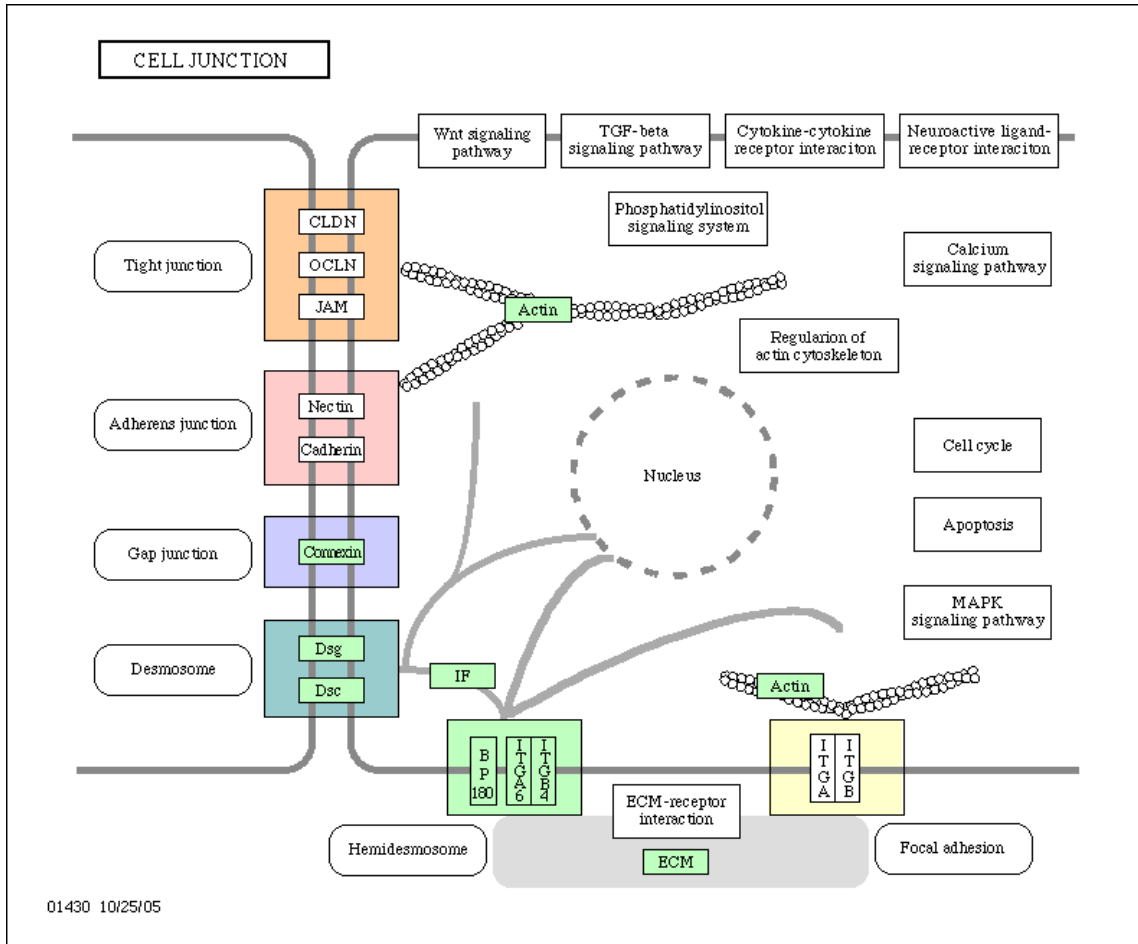


Figure 2: Cell communication pathway. Result from the KEGG Pathway analysis.



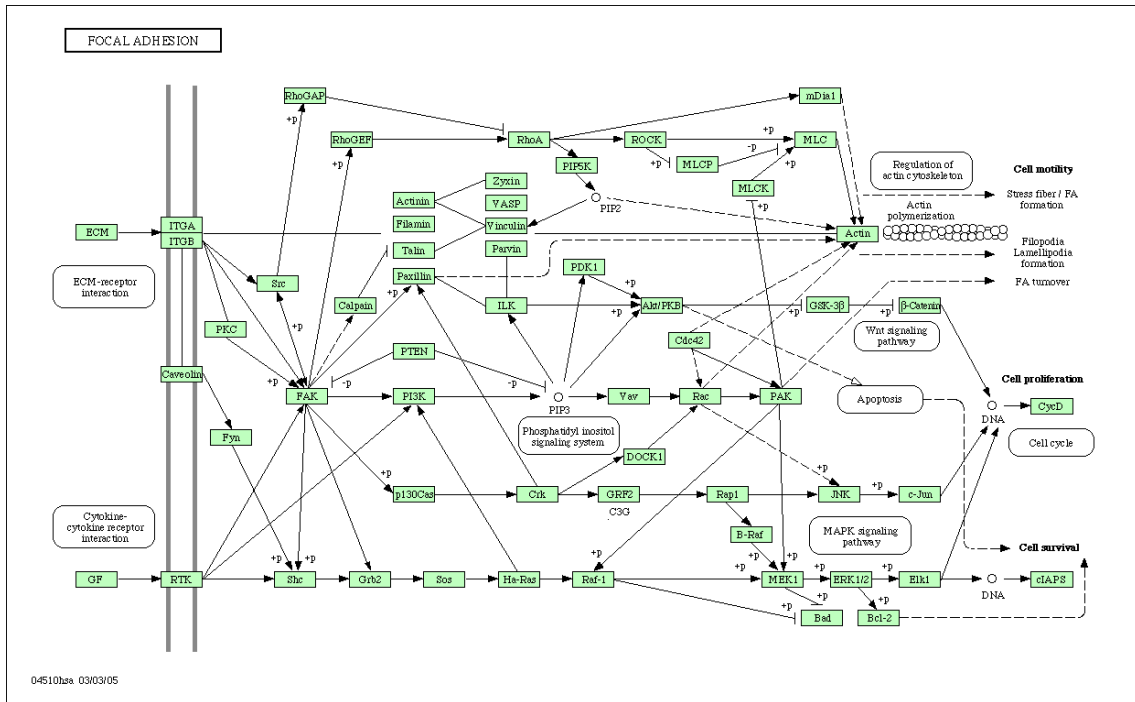


Figure 3: Focal adhesion. Result from the KEGG pathway analysis.

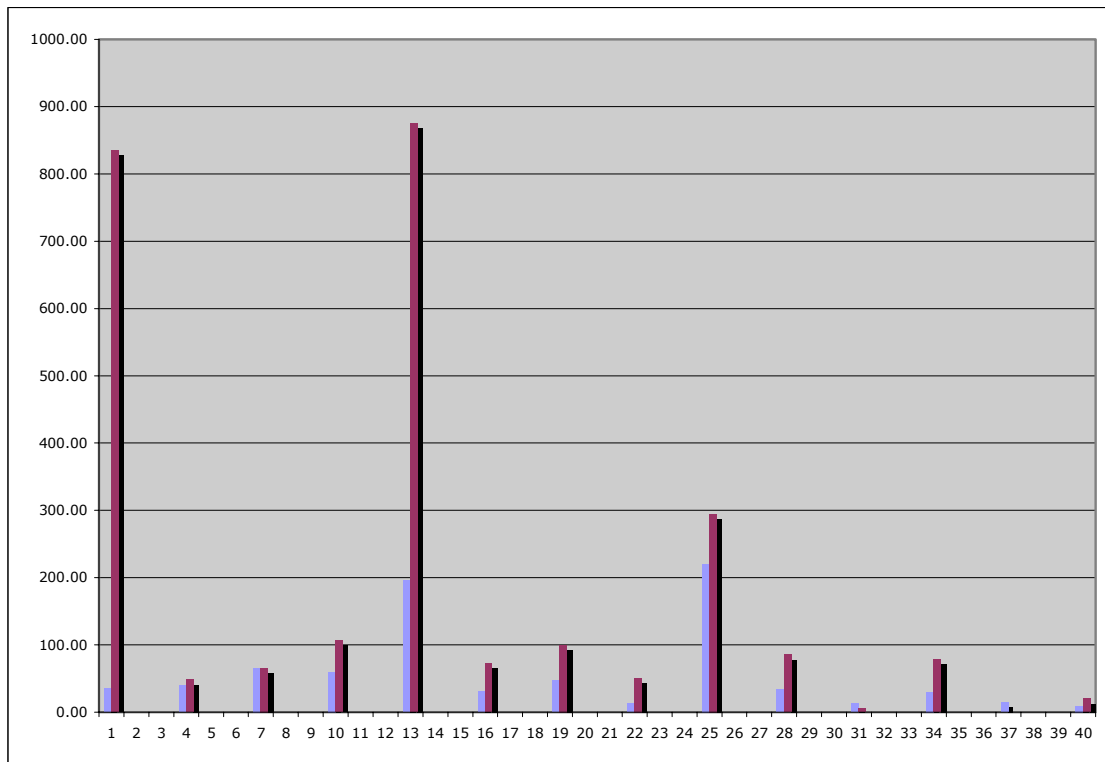


Figure 4: Results of RT-PCR for *MMP1*. The blue columns shows the RT-PCR quantity, while the red columns shows the corresponding microarray data for the tumour samples.

## References:

1. Ferlay J, Bray F, Pisani P, Parkin DM. Globocan 2002. Cancer incidence, mortality and prevalence worldwide.: IARC*Press*, 2004.
2. Gupta P, Warnakulasuriya S. Global epidemiology of areca nut usage. *Addiction Biology* 2002;7:77-83.
3. Chiba I. Prevention of betel quid chewers' oral cancer in the Asian-Pacific area. *Asian Pacific Journal of Cancer Prevention* 2001;2:263-9.
4. IARC. Betel-quid and areca-nut chewing and some areca-nut-derived nitrosamines.ed., vol. 85. Lyon: IARC, 2004.
5. Nair U, Bartsch H, Nair J. Alert for an epidemic of oral cancer due to use of the betel quid substitutes gutkha and pan masala: a review of agents and causative mechanisms. *Mutagenesis* 2004;19:251-62.
6. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nature Reviews* 2003;3.
7. Daftary DK, Murti PR, Bhonsle RB, Gupta PC, Mehta FS, Pindborg JJ. Oral squamous cell carcinoma. In: *Oral diseases in the Tropics.*, Prabhu, SR, Wilson, DF, Daftary, DK, Johnson, NW ed. Oxford: Oxford University Press, 1992.
8. CancerStats. Oral Cancer UK. Cancer Research UK, 2005.
9. Warnakulasuriya S, Ralhan R. Clinical pathological, cellular and molecular lesions caused by smokeless tobacco. *Journal of Oral Pathology and Medicine* 2006;(In press).
10. Scully C, Field JK, Tanzawa H. Genetic aberrations in oral or head and neck squamous cell carcinoma (SCCHN): 1. Carcinogen metabolism, DNA repair and cell cycle control. *Oral Oncology* 2000;36:256-63.
11. Reibel J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. *Crit Rev Oral Biol Med* 2003;14:47-62.
12. Nagpal JK, Das BR. Oral Cancer: reviewing the present understanding of its molecular mechanism and exploring the future directions for its effective management. *Oral Oncology* 2003;39:213-21.
13. Otero-Rey E, Garcia-Garcia A, Barros-Angueira F, Torres-Español M, Gándara-Rey JM, Somoza-Martin M. DNA microarrays in oral cancer. *Oral Medicine and Pathology* 2004;9:288-92.
14. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology in human cancer. *Oncogene* 2003;22:6497-507.
15. Harrington KJ, Nutting CM, Pandha HS. Gene therapy for head and neck cancer. *Cancer and Metastasis Reviews* 2005;24:147-64.
16. Modjtahedi. Molecular therapy of head and neck cancer. *Cancer and Metastasis Reviews* 2005;24:129-46.
17. Harris CC. Interindividual variation in binding of benzo[a]pyrene to DNA in cultured human bronchi. *Science* 1976;194:1067-069.
18. Cawson R, Eveson J. *Oral pathology and diagnosis*.: London: Gower, 1987.
19. Dysvik B, Jonassen I. J-Express: exploring gene expression data using Java. *Bioinformatics* 2001;17:369-70.
20. Efron B, Tibshirani R, Goss V, Chu G. *Microarrays and Their Use in a Comparative Experiment*. Stanford University, Division of Biostatistics, 2000.
21. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res.* 1996;6:986-94.
22. Choi P, Chen C. Genetic expression profiles and biological pathway alterations in head and neck squamous cell carcinoma. *Cancer* 2005;104:1113-128.

23. Cromer A, Carles A, Millon R, Ganguli G, Chalmel F, Lemaire F, Yuong J, Dembèlè D, Thibault C, Muller D, Poch O, Abecassis J, et al. Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene* 2004;23:2484-498.
24. O'Donnell RK, Kupferman M, Wei SJ, Singhal S, Weber R, O'Malley BJ, Cheng Y, Putt M, Feldman M, Ziober B, Muschel RJ. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene* 2005;24:1244-251.
25. Toruner GA, Ulger C, Alkan M, Galante AT, Rinaggio J, Wilk R, Tian B, Soteropoulos P, Hameed MR, Schwalb MN, Dermody JJ. Association between gene expression profile and tumor invasion in oral squamous cell carcinoma. *Cancer Genetics and Cytogenetics* 2004;154:27-35.
26. Scully C, Field JK, Tanzawa H. Genetic aberrations in oral or head and neck squamous cell carcinoma 2: chromosomal aberrations. *Oral Oncology* 2000;36:311-27.
27. Scully C, Field JK, Tanzawa H. Genetic aberrations in oral or head and neck squamous cell carcinoma 3: clinico-pathological applications. *Oral Oncology* 2000;36:404-13.
28. Stevens A, Lowe J. *Human Histology*. Elsevier Ltd., 2005.
29. Rajalalitha P, Vali S. Molecular pathogenesis of oral submucous fibrosis - a collagen metabolic disorder. *Journal Oral Pathology and Medicine* 2005;34:321-8.
30. 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, Wu L-W. The mRNA profile of genes in betel quid chewing oral cancer patients. *Oral Oncology* 2004;40:418-26.
31. Mesnil M. Connexins and cancer. *Biology of the Cell* 2002;94:493-500.
32. Jeon GA, Lee J-S, Patel V, Gutkind JS, Thorgeirsson SS, Kim EC, Chu I-S, Amornphimoltam P, Park MH. Global gene expression profiles of human head and neck squamous carcinoma cell lines. *International Journal of Cancer* 2004;112:249-58.
33. Franchi A, Santucci M, Sardi I, Paglierani M, Gallo O. Expression of matrix metalloproteinase 1, matrix metalloproteinase 2, and matrix metalloproteinase 9 in carcinoma of the head and neck. *Cancer* 2002;95:1902-910.
34. Thomas GT, Lewis MP, Speight PM. Matrix metalloproteinases and oral cancer. *Oral Oncology* 1999;35:227-33.
35. Nagata M, Fujita H, Ida H, Hoshina H, Inoue T, Seki Y, Ohnishi M, Ohyama T, Shingaki S, Kaji M, Saku T, Takagi R. Identification of potential biomarkers of lymph node metastasis in oral squamous cell carcinoma by cDNA microarray analysis. *International Journal of Cancer* 2003;106:683-89.
36. Alevizos I, Mahadevappa M, Zhang X, Ohyama H, Kohno Y, Posner M. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* 2001;20:6196-204.
37. Emberley E, Murphy L, Watson P. S100 proteins and their influence on pro-survival pathways in cancer. *Biochemistry and Cell Biology* 2004;82:508-15.
38. Wang G, Zhang S, Fernig D, Martin-Fernandez M, Rudland P, Barraclough R. Mutually antagonistic actions of S100A4 and S100A1 in normal and metastatic phenotypes. *Oncogene* 2005;24:1445-454.
39. Neville BW, Day TA. Oral cancer and precancerous lesions. *CA Cancer J Clin* 2002;52:195-215.
40. Koch BB, Trask DK, Hoffman HT, Karnell LH, Robinson RA, Zhen W, Menck HR. National survey of head and neck verrucous carcinoma. *Cancer* 2001;92:110-19.
41. Irie T, Aida T, Tachiwaka T. Gene expression profiling of oral squamous cell carcinoma using laser microdissection and cDNA microarray. *Med Electron Microsc* 2004;37:89-96.

42. Mèndez E, Cheng C, Farwell DG, Ricks S, Agoff SN, Futran ND, Weymuller J, E.A., Maronian NC, Zhao LP, Chen c. Transcriptional expression profiles of oral squamous cell carcinomas. *Cancer* 2002;95:1482-94.
43. Schliephake H. Prognostic relevance of molecular markers of oral cancer--A review. *International Journal of Oral and Maxillofacial Surgery* 2003;32:233-45.
44. Arora S, Matta A, Shukla NK, Deo SVS, Ralhan R. Identification of differentially expressed genes in oral squamous cell carcinoma. *Molecular Carcinogenesis* 2005;42:97-108.
45. Vigneswaran N, Wu J, Sacks P, Gilcrease M, Zacharias W. Microarray gene expression profiling of cell lines from primary and metastatic tongue squamous cell carcinoma: possible insights from emerging technology. *Journal of Oral Pathology and Medicine* 2005;34:77-86.
46. Warner GC, Reis PP, Jurisica I, Shilpi MS, Macmillan AC, Makitie AA, Grénman R, Reid N, Sukhai M, Freeman J, Gullane P, Irish J, et al. Molecular classification of oral cancer by cDNA microarrays identifies overexpressed genes correlated with nodal metastasis. *International Journal of Cancer* 2004;110:857-68.
47. Whipple ME, Mendez E, Farwell G, Agoff SN, Chen C. A genomic predictor of oral squamous cell carcinoma. *The Laryngoscope* 2004;114:1346-354.
48. Duffy MJ, Duggan C. The urokinase plasminogen activator system: a rich source of tumour markers for the individualised management of patients with cancer. *Clinical Biochemistry* 2004;37:541-8.
49. Hundsdorfer B, Zeilhofer H-F, Bock KP, Dettmar P, Schmitt M, Kolk A, Pautke C, Horch H-H. Tumour-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. *Journal of Cranio-Maxillofacial Surgery* 2005;33:191-6.
50. Tilakaratne WM, Klinikowski MF, Saku T, Peters TJ, Warnakulasuriya S. Oral submucous fibrosis: Review on aetiology and pathogenesis. *Oral Oncology*;In Press, Corrected Proof.
51. Chiu C-J, Chang M-L, Chiang C-P, Hahn L-J, Hsieh L-L, Chen C-J. Interaction of collagen-related genes and susceptibility to betel quid-induced oral submucous fibrosis. *Cancer Epidemiology, Biomarkers and Prevention* 2002;11:646-53.