

Manuscript 2

Gene Expression Profiles of Head and Neck Carcinomas from Sudanese and Norwegian Patients Reveal Common Biological Pathways Regardless of Race and Lifestyle

Bjarte Dysvik,¹ Endre N. Vasstrand,² Roger Løvlie,³ Osman A-Aziz Elgindi,⁸ Kenneth W. Kross,⁵ Hans J. Aarstad,⁵ Anne Chr. Johannessen,⁶ Inge Jonassen,^{1,7} and Salah O. Ibrahim⁴

Abstract Purpose: To explore possible range of gene expression profiles in head and neck squamous cell carcinomas (HNSCC) and pairwised normal controls from Sudanese ($n = 72$) and Norwegian ($n = 45$) patients using a 15K cDNA microarray and to correlate the findings with clinicopathologic variables.

Experimental Design: Samples from Sudan were grouped according to anatomic location/patients' habit of *toombak* (snuff) use, and 37 pools of 2 to 11 tumors matched to 37 pools of their normal controls from the same patients, respectively, were prepared. For Norway, eight pools of 3 to 11 tumors matched to eight pools of their normal controls from the same patients, respectively, were prepared according to anatomic location. Pools ($n = 45$) were hybridized to microarrays. For controls, 33 of the pools were hybridized against Human Reference RNA. Scanned array images were recorded, and data analysis was done in groups. For verification, results for selected genes were analyzed using quantitative real-time PCR/immunohistochemistry.

Results: We identified 136 genes from Sudan and 154 from Norway as differentially expressed between tumors and controls. Changes of the genes found were confirmed in >70% of the pools by hybridization against Reference RNA. Seventy-three genes and three main pathways (signal transduction, cell communication, and ligand-receptor interaction) were of relevance to the HNSCCs from both countries. Hierarchical clustering of the 73 genes identified subclasses of mixed tumors from the two populations, two independent subgroups for Norwegian tumors by their anatomic sites, and five subgroups for Sudanese tumors by their toombak habits. Quantitative real-time PCR/immunohistochemistry validated the microarray-based data.

Conclusions: Differences in gene expression between tumor and nontumor tissues were identified in HNSCCs. Analysis of the two population groups revealed a common set of 73 genes within three main biological pathways. This indicates that the development of HNSCCs is mediated by similar biological pathways regardless of differences related to race, ethnicity, lifestyle, and/or exposure to environmental carcinogens. Of particular interest, however, was the valuable association of gene expression signature found with toombak use and anatomic site of the tumors.

Head and neck squamous cell carcinomas (HNSCC) are among the 10 most prevalent cancers in the world, with a higher proportion occurring in developing countries (1–3). Tobacco use and/or alcohol consumption are the two principal risk factors involved in development of HNSCCs (1–3). In Norway, however, where these habits are common, incidence of HNSCCs

(1996–2001) has been found to be relatively low (6.12 per 100,000 per year for males and 2.51 per 100,000 per year for females; ref. 4). Contrary to this, in Sudan, where these habits are relatively low due to religious constraints, incidence of HNSCCs (1970–1985) is reported to be particularly high (11.60 per 100,000 per year for males and 6.91 per 100,000 per year for

Authors' Affiliations: Departments of ¹Informatics and ²Oral Sciences-Periodontology; ³Center for Medical Genetics and Molecular Medicine; ⁴Department of Biomedicine, Section of Biochemistry and Molecular Biology; ⁵Department of Surgical Sciences, Section of Otorhinolaryngology; and ⁶Department of Oral Sciences-Oral Pathology and Forensic Odontology, Haukeland University Hospital; ⁷Computational Biology Unit, Bergen Centre for Computational Science, University of Bergen, Bergen, Norway; and ⁸Department of Oral and Maxillofacial Surgery, Khartoum Dental Teaching Hospital, Sudan

Received 1/19/05; revised 11/19/05; accepted 12/2/05.

Grant support: Norwegian Research Council grant 155195/320/Salmon Genome Project, Norwegian Cancer Society grant 00053, Meltzer Høyskolefond, Richard With-Johnsen og Hustru Fanny's Fond, Fellelegatet til Fordel for Biologisk Forskning, FUGE technology platforms for microarrays (Norwegian Microarray Consortium) and for bioinformatics (Computational Biology Unit), and Unio

Internationale Contra Cancrum International Cancer Technology Transfer Fellowship (S.O. Ibrahim).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data and microarray data in MIAME compatible format for this article are available at <http://www.bioinfo.no/carcinoma>.

Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Requests for reprints: Salah O. Ibrahim, Department of Biomedicine, Section of Biochemistry and Molecular Biology, University of Bergen, Jonas Lies Vei 91, 5009 Bergen, Norway. Phone: 47-55-586-423; Fax: 47-55-586-360; E-mail: sosman@biomed.uib.no.
© 2006 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-05-0115

females; ref. 5). The high incidence observed in Sudan is attributed to wide use of smokeless tobacco, as oral snuff, locally called *toombak*, a mixture of tobacco powder and sodium bicarbonate (6). Despite improvements in surgery, chemotherapy, and radiation therapy, the overall survival associated with HNSCCs has not improved in the past 30 years (7).

HNSCCs result from progressive accumulation of genetic lesions with a precise nature still largely unknown (8–10). Identification of genes involved in carcinogenesis of HNSCCs may provide key factors to understanding these tumors and might lead to development of diagnostic markers and/or effective therapeutic strategies (8–10). Genetic alterations in HNSCCs result in qualitative and quantitative changes in gene expression profile, leading to abnormal cell function and proliferation (8–10). Microarray technology is a powerful tool that allows for studying the molecular basis of interactions on a large scale difficult to obtain through conventional methods of analysis (11–13). The technology is now widely used for disease diagnostics (14, 15), candidate gene identification (16, 17), gene expression profiling (18, 19), and elucidation of biological pathways (20–22). In the West, several reports have been published on microarray studies of HNSCCs with promising findings (reviewed in refs. 23–26). However, although the burden of HNSCCs is high in developing countries, studies using microarrays are rare, which might be attributed to the high cost of the technology. A comparison between HNSCCs from developing countries and developed ones using microarrays has not been attempted. We therefore wanted to explore possible range of gene expression profile differences and/or similarities between samples of HNSCCs and pairwise normal oral mucosal tissues (NOMT) from Sudanese ($n = 72$) and Norwegian ($n = 45$) patients, and to correlate the findings with clinicopathologic variables using the human cDNA DNR Microarray 15K arrays manufactured and supplied by Norwegian Microarray Consortium (<http://www.mikromatrise.no>).

Materials and Methods

Patient characteristics and biopsy specimens. The study protocol, adhering to acceptable standards for each country, was approved by Committees for Medical Ethics at Haukeland University Hospital, Norway, and University of Khartoum, Sudan. An identical protocol was used for tissue collection/processing in both countries. From December 1999 to July 2003, primary HNSCCs ($n = 117$; 72 from Sudan and 45 from Norway) with pairwise NOMTs ($n = 117$) were obtained from patients who had undergone surgery at the Department of Surgical Sciences, Section of Otorhinolaryngology, Haukeland Hospital and Department of Oral and Maxillofacial Surgery, Khartoum University Dental Teaching Hospital. NOMTs were obtained either from the contralateral site to that of the tumor or from normal mucosa that was at least 4 to 5 cm away from the tumor and was macroscopically normal. After surgery, samples were immediately submerged in RNAlater (Ambion, Austin, TX) and stored at -20°C . All samples were dispatched to Department of Biomedicine, University of Bergen, on dry ice and stored at -20°C until RNA purification and array experiments.

Using H&E-stained sections from snap-frozen and/or 10% formalin-fixed tissue blocks, all tumors were histopathologically confirmed as SCC by two of the authors (A.C.J. and S.O.I.) and staged by the 1987 Unio Internationale Contra Cancrum staging system. Tumors were pathologically graded into high (grade 1), moderate (grade 2), or poorly (grade 3/4) differentiated, according to Cawson and Eveson (27). To rule out stromal cell contamination, each tumor used here was confirmed histopathologically to contain $\geq 70\%$ tumor tissues

Table 1. Patient's clinicopathologic variables in the HNSCCs examined from patients from Sudan and Norway

| Clinical feature | Sudan samples, n (%) | Norway samples, n (%) |
|--------------------------------|---------------------------|----------------------------|
| Total no. samples | 72 | 45 |
| Gender | | |
| Male, n (%) | 48 (67) | 31 (69) |
| Female, n (%) | 24 (33) | 14 (31) |
| Median age (y) | 60 | 65 |
| Tobacco use*, n (%) | 17 (24) | 32 (71) |
| Alcohol use†, n (%) | 21 (29) | 27 (60) |
| Snuff dipping‡, n (%) | 30 (42) | ND§ |
| Anatomic location, n (%) | | |
| Oral cavity | 72 (100) | 32 (71) |
| Oropharynx | 0 | 5 (11) |
| Larynx | 0 | 6 (13) |
| Sinus | 0 | 2 (4) |
| Clinical stage, n (%) | | |
| Stage I-II | 31 (43) | 23 (51) |
| Stage III-IV | 41 (57) | 22 (49) |
| Tumor differentiation, n (%) | | |
| Well | 34 (47) | 16 (36) |
| Moderate | 24 (33) | 22 (49) |
| Poor | 14 (19) | 7 (16) |

Abbreviation: ND, not determined.

*Current or past history of tobacco use. Smoking was quantified by number of cigarettes smoked per day, and accordingly, patients were grouped into nonsmokers; light smokers (1-10 cigarettes per day), moderate smokers (11-20 cigarettes per day), and heavy smokers (> 20 cigarettes per day).

†For alcohol drinking, patients were grouped into nondrinkers, light drinkers (1-2 drinks per day), moderate drinkers (3-5 drinks per day), and heavy drinkers (> 5 drinks per day).

‡Data indicates users of the Sudanese snuff, toombak.

§Data not determined due to lower association of this type of Scandinavian snuff with HNSCCs in Norway.

and $< 10\%$ necrotic debris by analysis of corresponding H&E sections. Patients' data on clinicopathologic variables, smoking, alcohol drinking, and/or snuff dipping were obtained (either as part of routine clinical examination/history taking or from hospital records) and presented in Table 1 (Supplementary Table S1 at <http://www.bioinfo.no/carcinoma>).

RNA extraction/cDNA synthesis, labeling, and hybridizations. The arrays used were printed on CMT-GAPS slides with 50% DMSO-printing solution, containing a set of 15,000 sequenced-verified human cDNA clones selected from Research Genetics UniGene (I.M.A.G.E) 40K human clone collection set representing both known/unknown genes and expressed sequence tags. Total RNA was extracted using either TRIzol reagent (Life Technologies/Bethesda Research Laboratories, Gaithersburg, MD) and/or RNeasy Fibrous Tissue kit (Qiagen, Chatsworth, CA), according to the manufacturer's instructions. RNA quantity and quality were evaluated using spectrophotometer, 2% agarose gel electrophoresis, and minichips (Bioanalyzer 2100, Agilent Technologies, Wilmington, DE). Following RNA extraction/quantification, and due to lack of sufficient RNA from the majority of samples to perform separate hybridizations, we combined and pooled in equal quantities RNA samples from corresponding tumors and controls (as biological averaging), by grouping of samples according to country of origin, anatomic location, and patients' habit of toombak (Sudan cases) use. Accordingly, 37 pools (36 pools of two patients each and 1 pool of

11 patients together) of RNA were prepared and grouped for Sudan samples, and eight pools (7 pools of three to eight patients each and 1 pool of 11 patients together) were grouped for Norway samples. To control for array experiments and to rule out for field cancerization effect observed in HNSCCs when using normal mucosa adjacent to resected tumors as controls, 33 of the pools prepared (29 from tumors/4 from controls) were used for array analysis by hybridization against Stratagene Universal Human Reference RNA. Details of study design for all the hybridizations done are illustrated at <http://www.bioinfo.no/carcinoma>. RNA (20-25 μ g) labeling was done using FairPlay Microarray Labeling Kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. Labeled cDNAs were hybridized at 45°C to microarrays in the Ventana Discovery (Ventana Medical Systems, Strasbourg, France) station according to the manufacturer's protocols. Hybridized slides were washed and scanned (5-10 μ m resolution) on a GenePix 4000B scanner (Axon Instruments, Union City, CA). Summary of the protocols used for RNA extraction, cDNA synthesis, and labeling/hybridizations are available at the BioArray Software Environment at <http://www.mikromatrise.no>.

Image quantification, data collection, and statistical analysis. Scanned TIFF images from GenePix scanner for Cy5 and Cy3 were analyzed using GenePix Pro 5.0 software (Axon Instruments). Image quantitation files obtained for all of the targets, related statistics, and merged jpg image file for both channels were imported into J-Express data analysis software (ref. 28; <http://www.molmine.com/jexpress>), where data from each array were subjected to filtering/normalization (see below). The spotpix suite in J-Express was used for quality control and low-level processing of each array. After data preparation, a log₂ ratio of median background corrected signals for each channel was calculated and inserted into an expression matrix.

Arrays used were from different print runs but with the same set of 15,000 genes. In total, 100 experiments (including two to three times rehybridization for some cases) were carried out using eight batches of microarrays (see schematic list of hybridizations at <http://www.bioinfo.no/carcinoma>). This was considered in the analysis procedure, and data from each experiment were assigned to corresponding array batch generating corresponding groups of data set with two experimental designs. The first one corresponds to hybridizations of RNA pools from tumors versus RNA pools of corresponding controls. The second one corresponds to hybridization of RNA pools from tumors and controls versus Reference RNA. Furthermore, to reduce variance from array/batch bias, an expression matrix was created for each group and analyzed separately.

To extract candidate genes from the microarrays, separate statistical analysis of the data was done. For each array, data were filtered, and spots satisfying at least one of the following three criteria were discarded: (a) spots flagged by GenePix; (b) spots with <20% of its foreground pixels being >2 SDs above background intensity in both channels; and (c) spots being empty or spikes/controls followed by applying a global lowess

normalization method (29). Accordingly, genes with >50% of their values filtered (i.e., missing) were removed, and missing values were set to zero in the data set groups 1 to 4/6 to 8 (Fig. 1), indicating equal expression in tumors/controls. In group 5 (see Fig. 1), missing values were imputed using KNN impute method with $k = 10$ (30). To identify differentially expressed genes for groups 1 to 4/6 to 8, two synthetic perfect candidate expression profiles were built (one for up-regulated and one for down-regulated genes; see Fig. 1), and all gene expression profiles were sorted with respect to Euclidean distance from this candidate. For group 5, the t score was evaluated for each gene between the two groups of tumors and normals versus Reference RNA using the feature subset selection functionality in J-Express. To avoid false positives in each defined list, a threshold value was set manually so that genes included showed clear differential expression in the corresponding data set for each group (judged by visual inspection), and accordingly, 35 to 578 genes were found for the eight expression matrices. Furthermore, we searched for genes present in several of these lists and sorted them according to their number in the lists they occurred in. The statistical significance of the results was then assessed using a permutation study where eight lists of the same length were randomly generated by sampling from the filtered sets of genes for each data set. Accordingly, 100 random samplings were done, and the average number of genes occurring in one, two, three lists, and so on was found. This resulted in an average occurrence of 0.27 genes in five lists and of 2.08 in four lists. For comparison, obtained lists of genes were further studied, and three genes were found to occur in five of the lists, whereas 14 were found to occur in four of the lists. This ranking procedure was used, and analysis was continued to search for most significant top discriminating genes.

Finally, and to determine global similarities/comparability in gene expression profile for the data across the HNSCCs examined, an unsupervised two-way hierarchical cluster analysis of the candidate genes was done with J-Express.

Quantitative real-time PCR. For an independent validation of the array data, we selected four of the genes (*keratin 4*, *S100A2*, *Ku-70*, and *fibronectin*) found overrepresented in the eight candidate lists and examined their expression levels by quantitative real-time PCR (Taqman) using aliquots (200-300 ng) of RNA [from 29 pairs of tumors/one pool containing all normals (same RNA used for arrays)/RNA from individual 69 tumors/3 normals] that were reverse-transcribed in a 50- μ L reaction volume using High-Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA) following the manufacturer's protocol. Real-time PCR reaction was carried out in 384-well microtiter plates on ABI Prism Sequence Detector 7900 HT (Applied Biosystems) with Assay-on-Demand kit (Applied Biosystems). Probe sequences used for PCR for *keratin 4* (Hs00361611_m1), *S100A2* (Hs00195582_m1), *Ku-70* (Hs00750856_s1), *fibronectin* (Hs00415006_m1), and endogenous control β -*actin* (Hs99999903_m1) are available on request from the author. For each sample of each gene, PCR amplification was done in triplicate with the endogenous control.

Fig. 1. Assignment of data from each experiment to corresponding batch/experimental design generating eight groups of data set with two experimental designs. Groups 1-4/6-8 correspond to hybridization of total RNA from tumors versus total RNA from normal controls. Group 5 corresponds to hybridization of total RNA from tumors and normal controls versus total RNA from the Universal Human Reference.

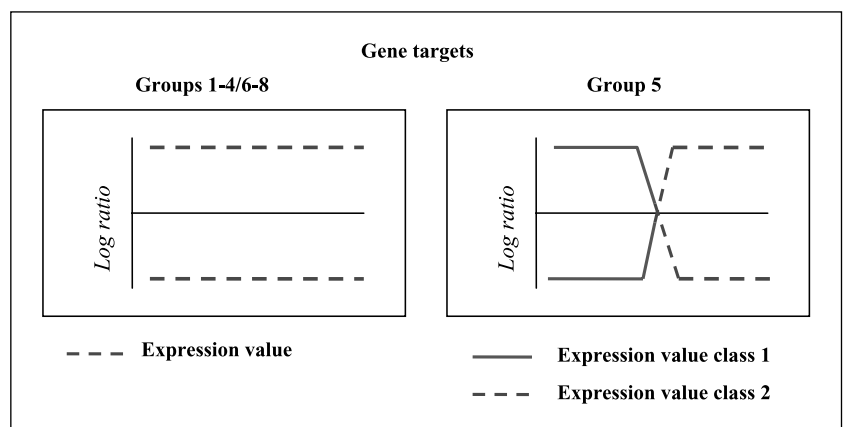


Table 2. The 73 genes that showed up-regulation or down-regulation in the HNSCCs examined from patients from both Sudan and Norway

| Accession no. | Gene symbol | Cluster | Gene name/or description | Average fold change |
|---------------|----------------------|----------------------|--|---------------------|
| 842802 | <i>G22P1</i> | Hs.292493 | Thyroid autoantigen 70 kDa (Ku antigen) | -4.76 |
| 729942 | <i>SPRR2C</i> | Hs.2421 | Small proline-rich protein 2C | 1.09 |
| 415388 | <i>ACYP1</i> | Hs.18573 | Acylphosphatase 1, erythrocyte (common) type | -1.35 |
| 39885 | <i>DBCCR1L</i> | Hs.65765 | Similar to RIKEN cDNA B830045N13, DBCCR1-like | -5.06 |
| 810813 | <i>S100A2</i> | Hs.516484 | S100 calcium binding protein A2 | 1.84 |
| 139009 | <i>FN1</i> | Hs.203717 | Fibronectin 1 | 1.24 |
| 1035889 | <i>KRT4</i> | Hs.371139 | Keratin 4 | -1.75 |
| 592540 | <i>KRT5</i> | Hs.433845 | Keratin 5 | 1.85 |
| 877827 | <i>RPS27A</i> | Hs.311640 | Ribosomal protein S27a | -1.17 |
| 813742 | | In multiple clusters | PTK7 protein tyrosine kinase 7, exportin 6 | 1.22 |
| 357892 | <i>NOTCH2NL</i> | Hs.502564 | Notch homologue 2 | -2.13 |
| 214565 | | In multiple clusters | Ribosomal protein S14 | -2.55 |
| 134476 | <i>SYBL1</i> | Hs.24167 | Synaptobrevin-like 1 | -2.16 |
| 122159 | <i>COL3A1</i> | Hs.443625 | Collagen, type III, alpha 1 | 1.49 |
| 23073 | <i>FGF2</i> | St. | Fibroblast growth factor 2 (basic) | 1.60 |
| 970590 | <i>MAPK1</i> | Hs.431850 | LOC150225 | 1.34 |
| 868368 | <i>TMSB4X</i> | Hs.522584 | Thymosin, beta 4, X chromosome | 1.15 |
| 813614 | <i>SPRR1B</i> | Hs.1076 | Small proline-rich protein 1B (cornifin) | 1.09 |
| 813823 | <i>LUM</i> | Hs.406475 | Lumican | 1.06 |
| 418126 | <i>U2AF1</i> | Hs.365116 | U2 (RNU2) small nuclear RNA auxiliary factor 1 | -2.53 |
| 1519515 | <i>HSH1N1</i> | Hs.270851 | HIV-1 induced protein Hin-1 | -2.25 |
| 782513 | <i>G1P3</i> | Hs.523847 | Interferon, alpha-inducible protein (clone IFI-6-16) | 1.31 |
| 50703 | | In multiple clusters | KIAA0169 protein, SH3-domain GRB2-like endophilin B2 | -9.13 |
| 550353 | <i>PIGF</i> | Hs.468415 | Phosphatidylinositol glycan, class F | -1.66 |
| 839991 | <i>COL1A2</i> | Hs.489142 | Collagen, type I, alpha 2 | 1.42 |
| 549146 | <i>TRIM22</i> | Hs.501778 | Tripartite motif-containing 22 | 1.48 |
| 307231 | <i>ADH1B</i> | Hs.4 | Alcohol dehydrogenase 1B (class I), beta polypeptide | 1.25 |
| 1502206 | <i>MGAM</i> | Hs.122785 | Maltase-glucoamylase (alpha-glucosidase) | -1.99 |
| 208718 | <i>ANXA1</i> | Hs.494173 | Annexin A1 | -1.09 |
| 204179 | <i>CARF</i> | Hs.32922 | Collaborates/cooperates with ARF protein | -3.20 |
| 418323 | <i>SDHA</i> | Hs.547025 | Similar to Succinate dehydrogenase flavoprotein subunit | -2.77 |
| 70692 | <i>SERPINB2</i> | Hs.514913 | Serine (or cysteine) proteinase inhibitor, clade B, member 2 | -1.39 |
| 771016 | <i>PTPN18</i> | Hs.516390 | Protein tyrosine phosphatase, non-receptor type 18 | -2.03 |
| 81331 | | In multiple clusters | With strong similarity to protein pir:I56326 (<i>H. sapiens</i>) | 1.26 |
| 210646 | <i>PRKAG2</i> | Hs.131133 | Protein kinase, gamma 2 noncatalytic subunit | -1.56 |
| 810734 | <i>POLD4</i> | Hs.523829 | Polymerase (DNA-directed), delta 4 | 1.59 |
| 704290 | <i>ECE1</i> | Hs.195080 | Endothelin converting enzyme 1 | 1.29 |
| 1897302 | | Not in UniGene | Serine/threonine kinase 17b (apoptosis inducing) | -2.22 |
| 203514 | <i>KIAA1991</i> | Hs.369437 | Hypothetical protein FLJ00225 | -1.56 |
| 796198 | <i>EFNB2</i> | Hs.149239 | Ephrin-B2 | 1.16 |
| 2027587 | <i>CD19</i> | Hs.96023 | CD19 antigen | -1.69 |
| 501890 | <i>cyclophilin A</i> | Hs.505141 | similar to Peptidylprolyl isomerase A, isoform 1 | -1.53 |
| 258026 | | Hs.174273 | <i>Homo sapiens</i> - transcribed sequences | -1.30 |
| 68894 | <i>FCGRT</i> | Hs.111903 | Fc fragment of IgG, receptor, transporter, alpha | -1.60 |
| 250654 | <i>SPARC</i> | Hs.111779 | Secreted protein, acidic, cysteine-rich (osteonectin) | 1.26 |
| 782793 | | Hs.535454 | <i>Homo sapiens</i> , similar to LOC161981 | 1.31 |
| 183120 | <i>CRSP6</i> | Hs.444931 | <i>Homo sapiens</i> - transcribed sequences | -2.55 |
| 220096 | <i>CNGB1</i> | Hs.147062 | Cyclic nucleotide-gated channel beta 1 | -1.21 |
| 2306221 | <i>WNT10B</i> | Hs.91985 | Wingless-type MMTV integration site family, member 10B | -1.77 |
| 1638971 | <i>UBE3A</i> | Hs.22543 | SNRPN upstream reading frame | -1.51 |
| 202194 | <i>SPTA1</i> | Hs.119825 | Spectrin, alpha, erythrocytic 1 (elliptocytosis 2) | -1.62 |
| 768561 | <i>CCL2</i> | Hs.303649 | Chemokine (C-C motif) ligand 2 | 1.13 |

(Continued on the following page)

Table 2. The 73 genes that showed up-regulation or down-regulation in the HNSCCs examined from patients from both Sudan and Norway (Cont'd)

| Accession no. | Gene symbol | Cluster | Gene name/or description | Average fold change |
|---------------|---------------|----------------------|---|---------------------|
| 242070 | | Hs.16222 | <i>Homo sapiens</i> – transcribed sequences | –1.42 |
| 68049 | <i>IGFBP5</i> | Hs.369982 | Insulin-like growth factor binding protein 5 | 1.18 |
| 628418 | <i>GPD1</i> | Hs.524418 | Glycerol-3-phosphate dehydrogenase 1 (soluble) | –1.52 |
| 245277 | <i>ITM2A</i> | Hs.17109 | Integral membrane protein 2A | 1.25 |
| 487820 | <i>CEBPD</i> | Hs.440829 | CCAAT/enhancer binding protein (C/EBP), delta | –1.57 |
| 487959 | | Not in UniGene | Eleven nineteen lysine-rich leukemia gene-associated factor 1 | –1.82 |
| 207771 | | Hs.37648 | Sapiens cDNA FLJ31407 fis, clone NT2NE2000137 | –1.39 |
| 293539 | | Hs.269113 | With weak similarity to protein sp:P11369 | 1.32 |
| 810358 | | In multiple clusters | Acyl-Coenzyme A dehydrogenase | –1.01 |
| 1552481 | <i>LCP2</i> | Hs.304475 | Lymphocyte cytosolic protein 2 | –1.37 |
| 646556 | <i>ZFP85</i> | Hs.427120 | Moderately similar to zinc finger protein 85 | –1.73 |
| 1016051 | <i>ROCK1</i> | Hs.306307 | Rho-associated, coiled-coil containing protein kinase 1 | –2.07 |
| 1543395 | <i>CCRK</i> | Hs.522274 | Cell cycle – related kinase | 2.19 |
| DNR98 | | Not in UniGene | DKF2-p418-K2134 | –7.26 |
| DNR110 | | Not in UniGene | 418G-1157-Coas2 | –2.25 |
| DNR181 | | Not in UniGene | DKF2-p418-Do116 | –3.27 |
| DNR212 | | Not in UniGene | OS bibl-12 | –3.41 |
| DNR232 | | Not in UniGene | OS bibl-3 | –2.06 |
| DNR240 | | Not in UniGene | ATCC 37940 Osteonectin | 1.34 |
| DNR288 | | Not in UniGene | Leptin | –1.92 |
| DNR388 | | Not in UniGene | Collagenase 3 | –1.29 |

NOTE: DNR Refers to the Norwegian Radium Hospital. These are sequences of known genes that have been included in the microarrays.

Immunohistochemistry. Expression of the same four genes examined by real-time PCR was validated by immunohistochemistry in corresponding paraffin-embedded tissues of 47 HNSCCs (27 from Sudan and 20 from Norway) and 11 NOMTs (from Sudan) analyzed in microarrays. Sections (4–5 μ m) were deparaffinized with xylene, dehydrated in graded ethanol, heated in a microwave for antigenic retrieval, incubated with DAKO peroxidase block 0.03% H₂O₂ containing sodium azide (Code K4007) for 5 minutes to eliminate endogenous peroxidase activity, washed in TBS for 10 minutes, and incubated with primary antibodies (60 minutes) on DAKO Autostainer Universal Staining System (DAKO, Copenhagen, Denmark) using antibodies against keratin 4 (Clone 6B10; dilution 1:20; Novocastra Laboratories, Newcastle upon Tyne, United Kingdom), S100A2 (A 5113; dilution 1:600, from DAKO), Ku-70 (A-9; dilution 1:500; Santa Cruz Biotechnology, Santa Cruz, CA), and fibronectin (F3648; dilution 1:400; Sigma-Aldrich, St. Louis, MO). After washing for 10 minutes in TBS, sections were incubated with Envision horseradish peroxidase (3,3'-diaminobenzidine) for 30 minutes, washed twice in TBS for 5 minutes each, and developed twice with 3,3'-diaminobenzidine-positive chromogen for 5 minutes each. Sections were counterstained with hematoxylin, rehydrated, and mounted using Eukitt. Cases in which primary antibody was omitted and substituted with TBS served as negative controls. Samples of oral SCCs known to show high expression of proteins examined were used as positive controls. Staining was graded as follows: –, no expression; +, <10% stained cells; ++, 10% to 50% stained cells; and +++, >50% stained cells. Gratings ++ and +++ were defined as overexpression.

Results

The study population included 79 males (48 Sudan and 31 Norway) and 38 females (24 Sudan and 14 Norway) with HNSCCs (median age of 60 years for Sudan and 65 years for Norway with male predominance of disease in both countries;

Table 1). Sudan cancers were located in the oral cavity (100%), whereas Norway cases were from oral (71%), larynx (13%), oropharynx (11%), and maxillary sinus (4%). Forty-one HNSCCs from Sudan and 22 from Norway were advanced stage tumors (stage III/IV). Among Sudanese cases, there were 30 (42%) snuff (toombak) dippers, 21 (29%) alcohol users, and 17 (24%) cigarette smokers. Thirty-two Norwegians (71%) smoked tobacco and 27 (60%) drank alcohol. Data on snuff use were not available for the majority of cases from Norway (Table 1; Supplementary Table S1). Although snuff is used widely in Norway, this is usually being either ignored or underestimated as a valuable clinicopathologic data recorded in HNSCCs.

Gene expression in tumors. Pairs of 45 pools of extracted RNA from 117 HNSCCs/117 NOMTs from patients from Sudan ($n = 72$) and Norway ($n = 45$) were studied by microarrays. Following data filtering/normalization, genes with significantly different expression patterns in tumors versus normals were analyzed in eight batches. Data from each experiment were assigned to corresponding batch/design. We identified 136 and 154 genes as differentially expressed between tumors and controls for the cases examined from Sudan and Norway, respectively. Genes detected were involved primarily in processes such as metabolism of carbohydrate, lipid, nucleotide and amino acids, genes involved in translation, signal transduction, cell communication and motility, and biodegradation of xenobiotics. Among the total number of genes ($n = 290$) found, 63 (22%) and 81 (28%) were differentially expressed between tumors and controls in the corresponding batches analyzed for cases from Sudan and Norway, respectively (Supplementary Tables S2a and 2b at <http://www.bioinfo.no/carcinoma>). Out of the same total number of genes, 73 (25%; 46 down-regulated

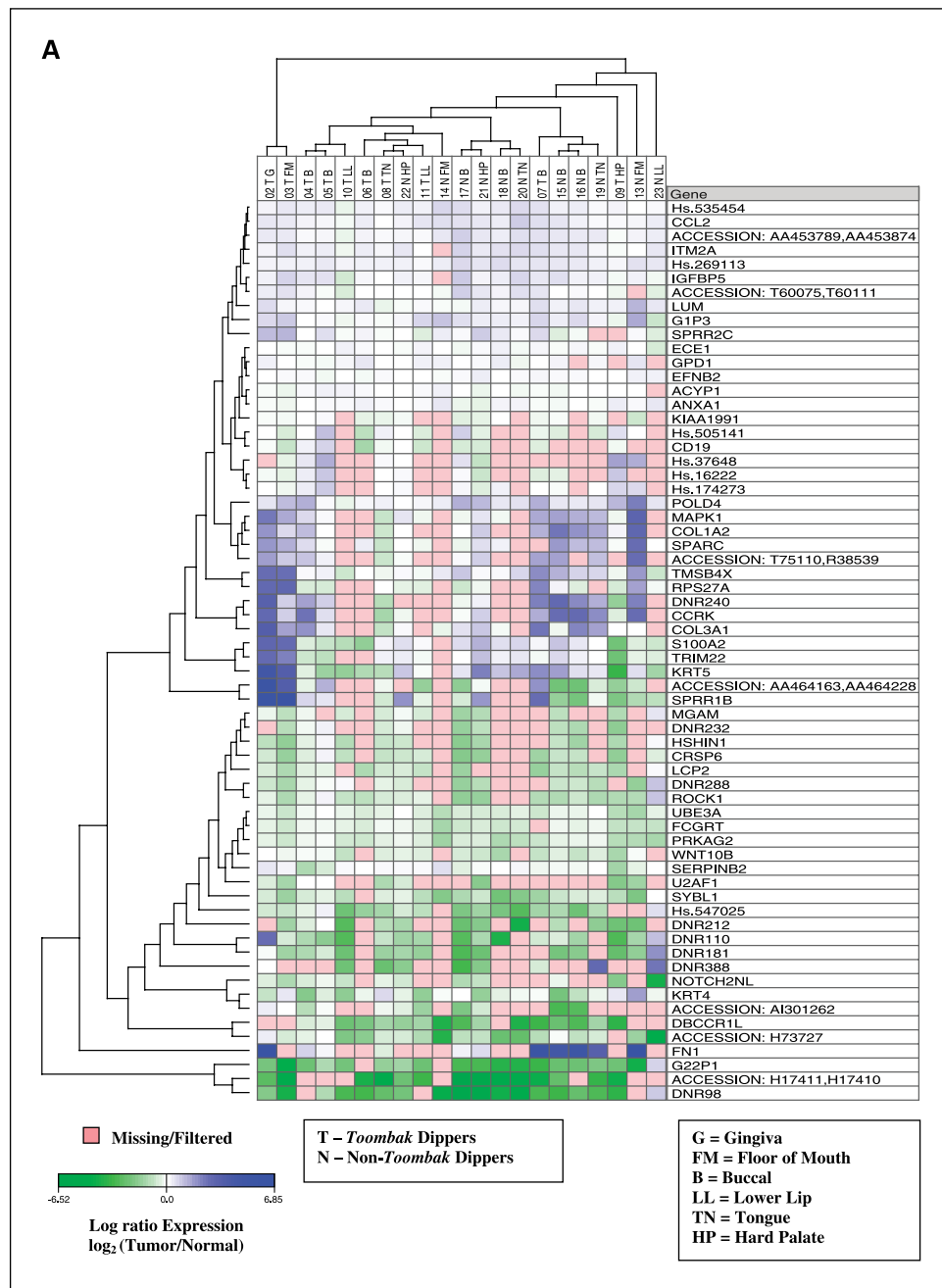


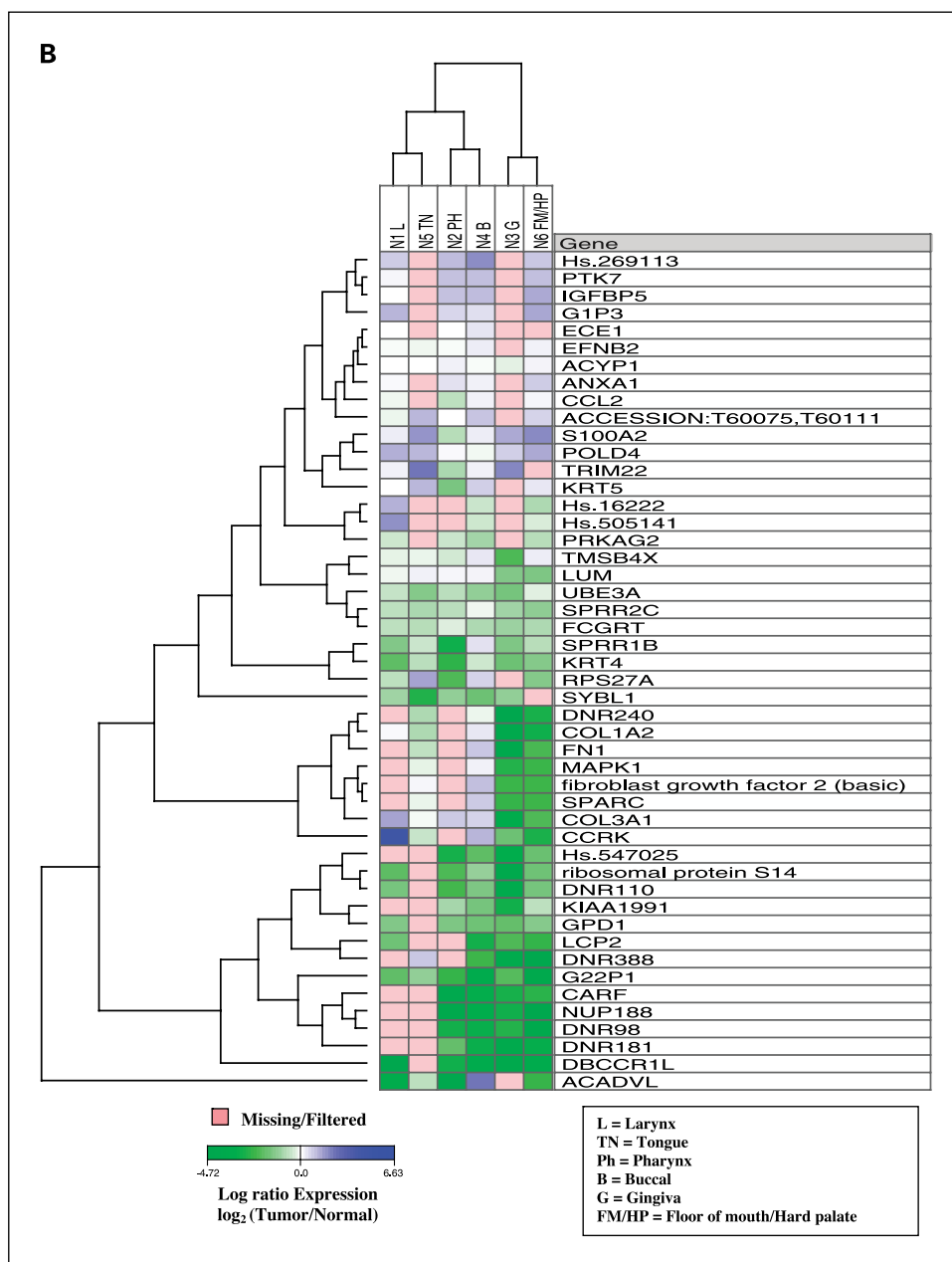
Fig. 2. A, an example of overall patterns of expression of the 73 genes found in tumor subclasses from patients from Sudan. Majority of tumors from Sudan tended to group tightly together on subclasses according to toombak use.

and 27 up-regulated; Table 2) were found expressed in tumors/controls for the cases from the two populations together. These included 53 (73%) known genes reported in HNSCCs. To obtain biological background knowledge of the 73 genes alone, and when including the 63 and 81 found in corresponding batches for each country alone, the eGOn (explore Gene Ontology; ref. 31; <http://nova2.idi.ntnu.no/egon/>) tool for gene functional classification was used. The 73 genes alone, and when included to the others, seemed under molecular function, biological process, and cellular component, respectively (Supplementary data at <http://www.bioinfo.no/carcinoma>).

We searched the Kyoto Encyclopedia of Genes and Genomes biochemical pathway database (ref. 32; <http://www.genome.jp/kegg/>) to determine altered/shared pathways in the cases examined. Of the 73 genes, we searched using 53 known genes and

found 24 associated pathways (Supplementary data at <http://www.bioinfo.no/carcinoma>). Furthermore, we used 46 and 62 known genes out of the 63 and 81 genes found only differentially expressed in cases from Sudan and Norway, respectively. The former was associated with 20 pathways and the latter with 14 (Supplementary data at <http://www.bioinfo.no/carcinoma>). Interestingly, three target pathways: signal transduction (wnt signaling), cell communication (focal adhesion), and ligand-receptor interaction (ECM-receptor interaction) were found to be common target pathways in the cases from the two countries (Supplementary data at <http://www.bioinfo.no/carcinoma>). When searching with the total number of known genes ($n = 161$) found, 43 associated pathways were found. Of interest, the same three pathways mentioned above besides cell communication/amino acid metabolism were found common in all cases.

Fig. 2 Continued. **B**, an example of overall patterns of expression of the 73 genes found in tumor subclasses from patients from Norway. For Norway, tumors formed groups based on anatomical sites.



Analysis of tumor groups and clinicopathologic data by clustering. We focused on the 73 genes to exclude dominance of genes with highly correlated expression profiles, which might dominate and define a cluster pattern, and subjected them to further analysis by clustering to determine gene expression global differences and/or similarities for individual tumors. Using unsupervised two-way clustering analysis on \log_2 ratio values, the results revealed variations in the differentially expressed genes among the tumors from the two countries with most of the cases showing a similar pattern of gene expression. Overall, tumors clustered into two main groups in patients from Sudan (Fig. 2A) and Norway (Fig. 2B) and had shown two major classes with mixed tumors from the two populations (Fig. 2C, *blue Arabic numerals* for Sudan cases and *red Arabic numerals* for Norway cases) as shown by the representative dendrograms for subgroups of the tumors.

Furthermore, we searched for association between separate classes of tumors from each country alone and clinicopathologic variables across all samples. We identified subclasses of five and two independent subgroups for tumors from Sudan (see Fig. 2A) and Norway (see Fig. 2B), respectively, and a mixed subgroup with tumors from the two populations (Fig. 2C). Sudan subgroups consisted of two toombak dippers (predominantly from buccal mucosa followed by lower lip, gingiva, floor of mouth, all of grade 2-3), nondippers, and two mixed (see Fig. 2A). The subgroups of nondippers were predominantly from buccal mucosa followed by tongue and hard palate (all of grade 1-3), whereas the two mixed were predominantly from buccal mucosa followed by tongue, floor of mouth, hard palate, and lower lip, and of grade 1 to 3. For Norway, tumors tended to form groups based on anatomic sites [larynx/tongue (predominantly of grade 1-2)/pharynx/buccal mucosa and gingiva/floor of

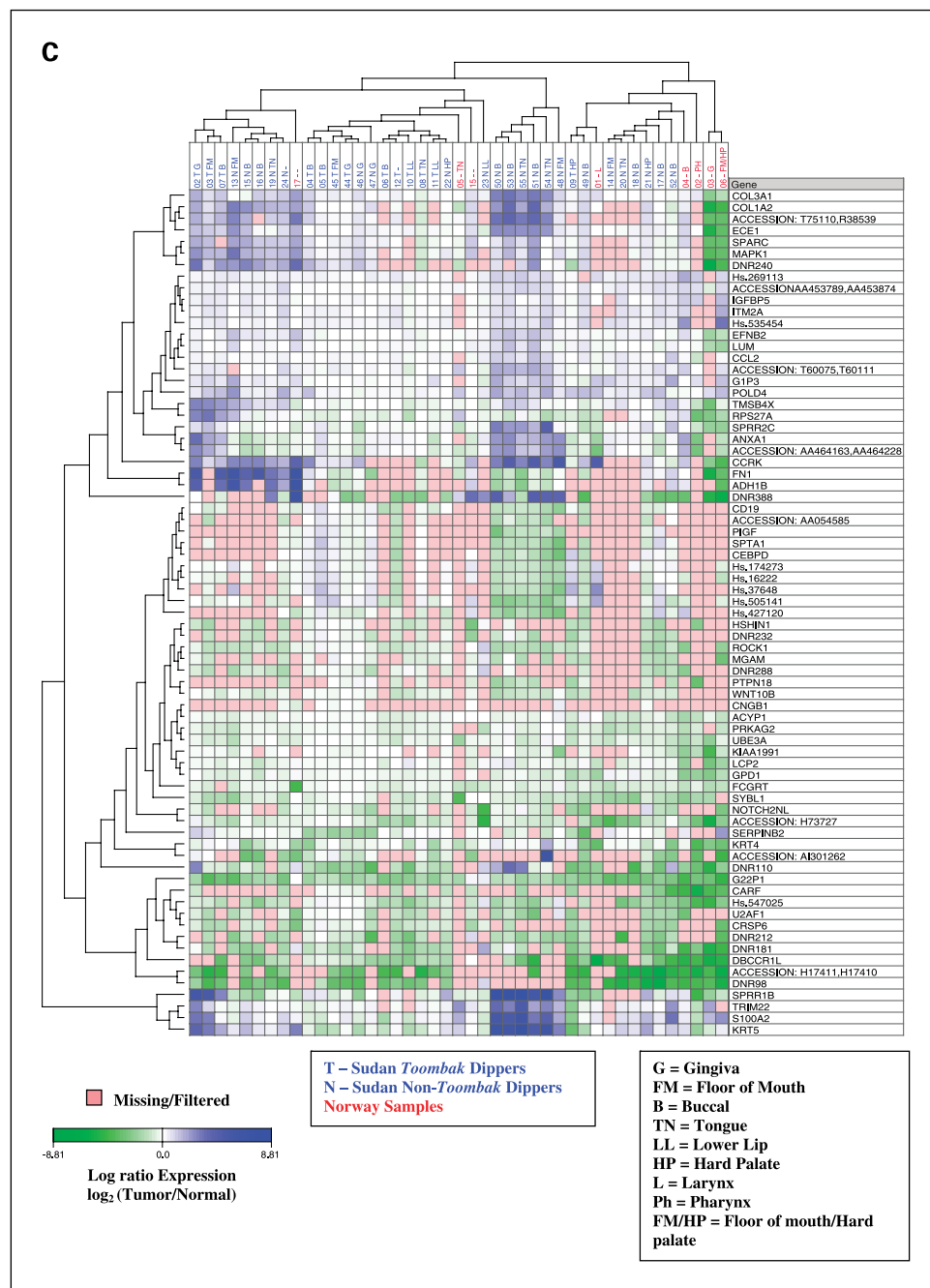


Fig. 2 Continued. C, an example of overall patterns of expression of the 73 genes found in tumor subclasses from patients from the two populations. Blue, overexpression in cancer cells; green, underexpression in cancer cells; white, unchanged expression (intensities of both Cy3 and Cy5 under the cutoff value). Graduated color patterns correspond to degree of expression changes. Tumors from Norway tended to share subclasses that were either clustered together with or close to subclasses of tumors from non – snuff dippers from Sudan, indicating that these tumors share a common expression pattern (C; blue Arabic numerals for Sudan cases and red Arabic numerals for Norway cases).

mouth/hard palate (predominantly of grade 3-4); see Fig. 2B]. When searching for relationships among the mixed subgroups of tumors from the two countries in terms of their similarities in expression profile by clinicopathologic variables/tobacco habits, we observed that the majority of the cases from Sudan tended to cluster together according to habit of toombak use (Fig. 2C, blue Arabic numerals). Most of the Norwegian cases (Fig. 2C, red Arabic numerals) tended to share subclasses that were either included or being close to subclasses of the non – snuff dipper (toombak) cases from Sudan, which might suggest a common expression pattern. However, when searching for more relationships using other clinicopathologic variables, no further associations were found. Although we found similarities between tumor groups, it is of note that the samples were collected in different decades

(1999–2003). Therefore, these findings need to be validated in careful experimental designs using larger sample size of tumors to be collected in different decades, which might provide important clues to the understanding of HNSCCs.

Comparison of genes identified with other HNSCC microarray findings. Because there are over 30 publications reporting on microarrays of HNSCCs, we did a comparison of our results (gene lists) to those reported in a subset of publications. First, we compared our list of 53 known genes (out of the 73) to Cancer Genome Anatomy Project (CGAP) Gene Library Summarizer database using the following queries: Library Group; CGAP libraries, Tissue Type; head and neck, Library Preparation; Any, Tissue Histology; Cancer, Library Protocol; Any. This resulted in a list of 2,177 genes involved in HNSCCs. By

comparing our gene list to this one, 15 genes (*COL1A2*, *COL3A1*, *FN1*, *KRT5*, *LUM*, *MAPK1*, *S100A2*, *SPARC*, *SPRR1B*, *TMSB4X*, and *TRIM22*, all up-regulated; and *KRT4*, *U2AF1*, *ANX1*, and *PIGF*, all down-regulated) were found as common. Thereafter, we selected a subset of nine published studies on microarrays of HNSCCs (33–41) by Medline database search and searched for our set of 53 genes in nine gene lists (containing 81, 102, 106, 138, 52, 153, 37, 115, and 7 genes, respectively) extracted from these studies. Because many genes have synonyms, we extracted all known synonyms for each gene in all the lists using the SOURCE unification tool (ref. 42; <http://source.stanford.edu/cgi-bin/source/sourceSearch>) and searched for all synonyms in our gene list. Thus, 10 of our 53 genes (with *CCL2*, *FIN1*, *GIP3*, *KRT5*, *MAPK1*, *S100A2*, *SPARC*, and *SPRR1B* as up-regulated and *UBE3A* and *KRT4* as down-regulated) were found in at least one of the nine lists. Thereafter, we compiled synonym lists for all genes ($n = 791$) reported in the nine studies and summarized the occurrence(s) for all synonyms to see how many genes were reported more than once. The results showed that 29 genes were common in two studies, and only four were common in three studies (Supplementary Table S2C at <http://www.bioinfo.no/carcinoma>), suggesting that the genetic signature for HNSCCs might be quite heterogeneous. Because it is possible that lesions anatomic locations may contribute to this, and that most of our cases pertain to cancers of the oral cavity/oropharynx, we excluded both sinus/laryngeal cancers in the Norwegian data set and reanalyzed our data to see if inclusion of these sites will affect gene expression found between tumors and controls examined from the two populations. However, and as expected, no changes were found in the final gene list(s) reported. Nevertheless, because HNSCCs includes cancers of the oral cavity, pharynx, and larynx, these findings demand analysis of larger sample size comprising tumors of more uniform characteristics that might provide important clues to the understanding of various gene networks implicated in HNSCC carcinogenesis.

We also analyzed the list of genes found in cases from Sudan and Norway, respectively, 63 and 81 genes, of which 46 and 62, respectively, are known (named) genes. We found that 7 of the 46 and 22 of the 62 genes were also in the CGAP (see Supplementary Tables S2a and b). This indicates that the majority of the known genes found only in cases from Norway were previously reported in the CGAPs, whereas those found in Sudan tumors were not. Although this might either be related to the fact that few studies were conducted in HNSCCs from developing countries or to the effect of toombak use, further analysis is required.

Quantitative reverse transcription-PCR/immunohistochemistry. To validate gene expressions found using independent methods, we used quantitative reverse transcription-PCR (RT-PCR)/immunohistochemistry, and the results were evaluated by three of the authors (S.O.I., A.C.J., and B.D.). With RT-PCR, we used triple determination and normalization based on β -actin level, and data were analyzed using GraphPad Prism software (GraphPad, San Diego, CA). A good correlation for the genes analyzed by RT-PCR (Fig. 3A-D; shown for the 29 pools of RNA from tumors/one pool from controls from Sudan/Norway) and corresponding protein expression [++ to +++ for overexpressed proteins (*S100A2* and fibronectin) and – to + for underexpressed protein (*Ku-70* and keratin 4) in most of the tumors/controls; Fig. 3E], was found. In addition, a good

correlation was found for the genes when analyzed by RT-PCR on separate individual RNA from 69 tumors/3 controls from the two countries (Fig. 3A-D; Supplementary data at <http://www.bioinfo.no/carcinoma>).

Discussion

We identified 136 genes from Sudanese and 154 from Norwegian patients with HNSCCs, as regulated differentially between tumors and normal controls. These genes encode proteins involved in signal transduction, cell communication, ligand-receptor interaction, cell communication, and amino acid metabolism pathways, which were shown to play an essential role in the development of HNSCCs (33–41, 43). The results found were controlled for in >70% of the cases by array analysis against Human Reference RNA. Although there are substantial differences in the initiating agents and risk factors related to the HNSCCs examined, 73 genes (46 down-regulated and 27 up-regulated) shared similar expression profile in the samples from the two countries, including 53 known genes (i.e., *Ku-70*, *keratin 4*, and *ANX1* found down-regulated and *S100A2*, *fibronectin 1*, *MAPK1*, *keratin 5*, *TRIM22*, *LUM*, *COL1A2*, *SPARC*, and *GIP3* found up-regulated, among others), previously reported in HNSCCs (33–41, 43, 44). In addition, three main pathways: Wnt signaling (signal transduction), focal adhesion (cell communication), and ECM-receptor (ligand receptor interaction) were found to be of relevance to the HNSCCs from the two populations. These findings add to the valuable microarray information available on HNSCCs, suggesting that this malignancy seems in part to be mediated by similar pathways regardless of differences in distinct etiology and/or other environmental risk factors. Results of some of the genes found here, like *trim22*, a member of the trim gene family, are also relevant to other reports on overexpression of another member of this family (*trim32*) in HNSCCs (43). Furthermore, and in a recent review on gene expression profiles involved in 24 HNSCCs studies, common gene expression alterations of some of the genes found in this study, like *fibronectin 1*, *keratin 4*, *ANX1*, *LUM*, *COL1A2*, *SPARC*, and *GIP3*, has also been reported in 11, 11, 7, 5, 10, 6, and 6 other HNSCC microarray studies, respectively (reviewed in ref. 44). Of note, however, some of the genes found in our work (i.e., up-regulated *small proline-rich protein 2C*, *ephrin-B2*, and *Notch homologue 2* and down-regulated *ribosomal protein S27a*, *synaptobrevin-like 1*, *cyclic nucleotide gated channel β 1*, *collagenase 3*, *coas2*, and *ROCK1*, among others) were not previously reported in HNSCCs, and their exact nature warrants further analysis. We used cDNA arrays in matched tissues of paired HNSCCs/NOMTs and pooled the samples due to lack of enough RNA for separate hybridizations. This approach, leading to biological averaging out of high and low gene expression, has helped to circumvent the problem of individual differences in gene expression and showed results that are in agreement with a subset of studies in HNSCCs (33–41) that either used fresh biopsy (33, 34), pure cells (35) or isolated cell cultures (36) using variety of array platforms, experimental designs, and methods of analysis.

In the literature, >30 microarray studies have been done in HNSCCs to determine gene expression changes during disease progression and/or to predict outcome of disease with similarity in genes identified/possible pathways involved

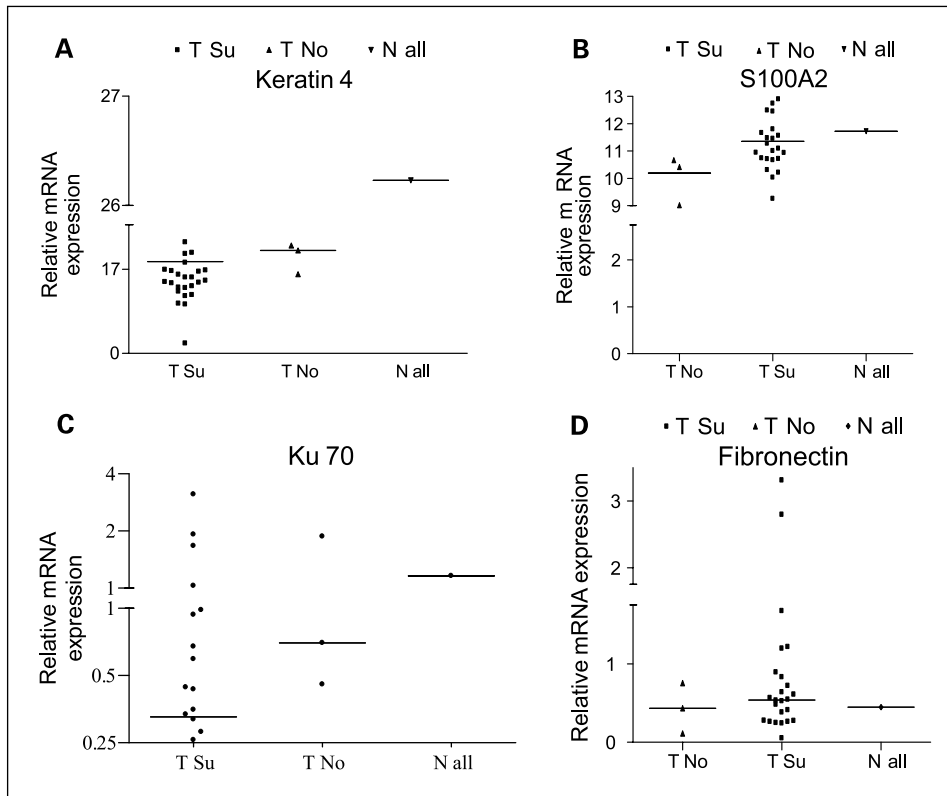


Fig. 3. Correlation for relative mRNA levels (fold change) for the four genes (*keratin 4*, *S100A2*, *Ku-70*, and *fibronectin*) analyzed (A-D) in pools of tumors (T) and all normal controls (N) from Sudan (Su) and Norway (No).

(reviewed in ref. 44). Nevertheless, probably due to limited number/type of samples (bulk/microdissected, primary cell cultures from biopsies), processing procedures, and diversity of technologies (low-density nylon membranes, high-density oligonucleotide chips), data from different studies vary widely, and it is difficult to assess the significance of the available findings (reviewed in ref. 44). Although direct comparison between these studies is rather difficult, we compared our findings (53 known genes) with nine published microarray studies on HNSCCs and with gene list in HNSCCs from the CGAP and searched for shared genes and/or pathways. Our identified genes targeted pathways, like MAPK signaling, cytokine-cytokine receptor interaction, regulation of actin cytoskeleton, adherens junction, ribosome, fatty acid metabolism, and DNA polymerase, among others, which are in agreement with subset of previous reports (33–41), and with other reports in HNSCCs (reviewed in ref. 44), suggesting that members of several pathways are altered in HNSCCs. In several of the listed studies (33–41), valuable information was achieved when comparing gene expression to patients' clinicopathologic variables. For example, Ginos et al. (33) found an association of gene expression signature enriched for genes involved in tumor invasion/metastasis with patients experiencing locally recurrent disease. Of note, the same tumors showed a marked absence of an immune response signature, suggesting that modulation of tumor-specific immune responses may play a role in local treatment failure (33). Belbin et al. (34) used clustering and found that molecular classification of 17 HNSCCs was a better predictor of disease-free survival than clinical/pathologic variables. Roepman et al. (37), using cDNA arrays, detected lymph node metastases for primary HNSCCs that arise in the oral cavity and oropharynx. We used clustering and searched for association between separate tumor classes and

clinicopathologic variables. We found independent subgroups of tumors from Sudan consisting of two toombak dippers, non-dippers, and two mixed. For Norway, tumors grouped based on different anatomic sites and were found to cluster with closely related results to tumors from non-toombak dippers from Sudan. Although in Sudan toombak is a recognized risk factor for oral cancer incidence, whereas in Norway, tobacco and alcohol are attributed to oral cancer development, it is difficult to eliminate the contribution of both factors to the gene expression found in either population. In addition, whereas our findings of similarities between the tumor groups studied might contribute substantially to the understanding of HNSCCs, it should be noted that decades in which the samples were collected were different. Therefore, these findings need to be validated in careful experimental designs, including analysis of larger sample size comprising tumor specimens to be collected within different decades and tumors available from databanks that have been collected and stored for several decades, which might provide important clues to the understanding of HNSCCs.

In this study, we searched the Kyoto Encyclopedia of Genes and Genomes database to find associated pathways using known genes found. Although many pathways were found, mainly three: signal transduction, cell communication, and ligand-receptor interaction showed commonality in Sudan and Norway tumors. Of interest, genes that dominated these pathways include down-regulated *ROCK1* and *ITGA4* and up-regulated *MMP-7*, *COL1A2*, *COL3A1*, and *fibronectin*. Although expression of these genes has been reported in HNSCCs (36–41), for *ROCK1* (called *ROKβ*, one of the best characterized member of the Rho GTPase family; ref. 45), this is the first report, which might offer new insight into a possible mechanism of Rho GTPase family in HNSCCs, albeit it remains

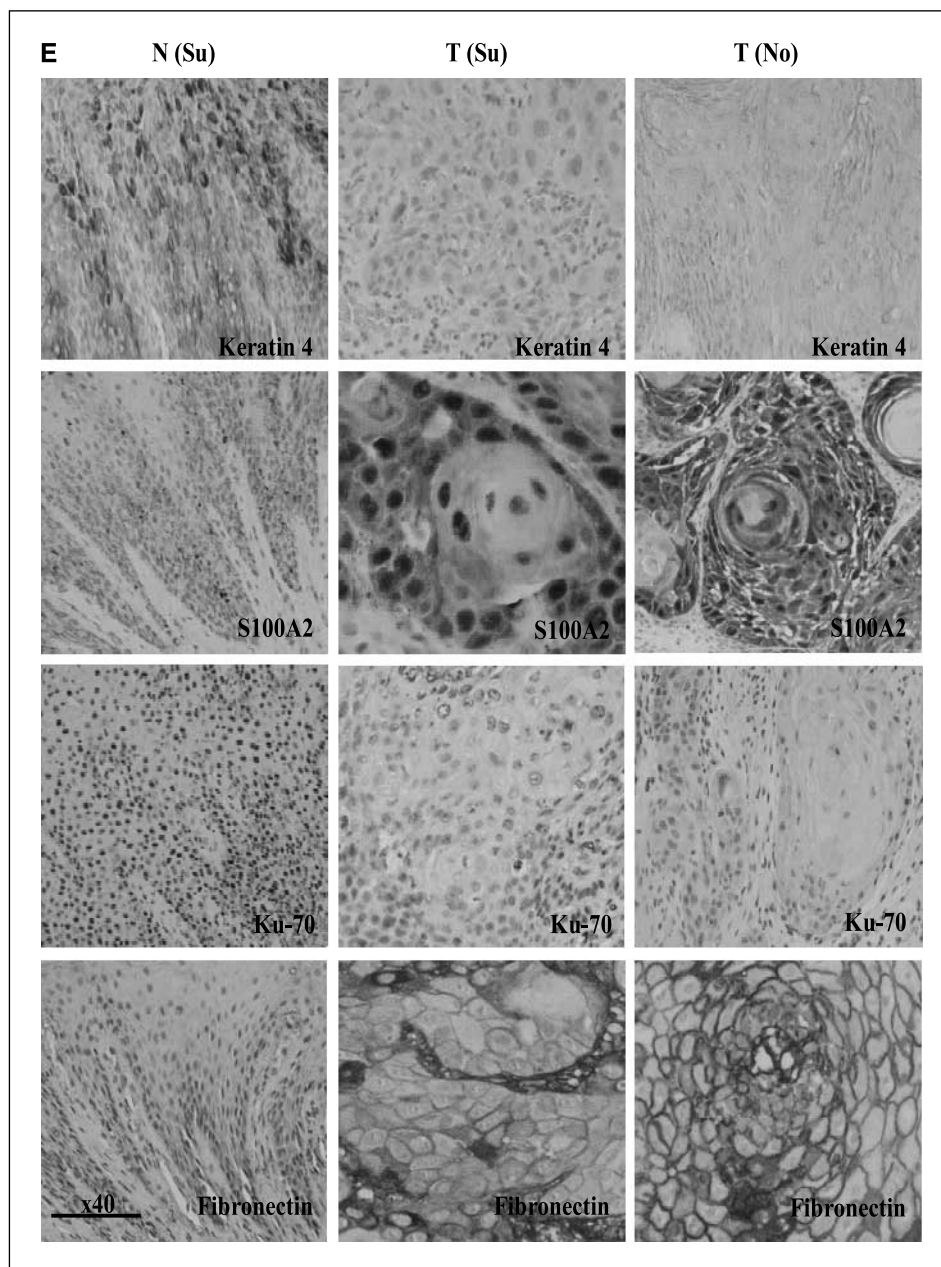


Fig. 3 Continued. *E*, corresponding protein expression (*E*) in representative formalin-fixed, paraffin-embedded tumor (*T*) and normal (*N*) samples from Sudan (*Su*) and Norway (*No*).

to be determined. *MMP-7*, the smallest of all the matrix metalloproteinases, degrades various matrix substrates, including proteoglycans, elastine, and gelatin, and cleaves nonmatrix proteins from the cell surface, including E-cadherin, pro-tumor necrosis factor- α , and Fas ligand (46). *COL1A2* and *COL3A1*, together with other types of collagens, participate in a variety of cellular processes, such as differentiation, tumorigenesis, and apoptosis, and loss of their expression has been reported in HNSCCs (47). Although variable loss or reduced expression of integrins has been reported in HNSCCs and because alterations in integrin expression in these tumors are variable, their role needs further investigations (47). By examining immunohistochemical expression of fibronectin in our cases, we found expression to occur throughout the stromal compartments associated with the tumors as shown by others (48). Thus, our data support the importance of the constituents of the cellular matrix interaction and their receptors in playing

prominent roles in the mechanisms involved in HNSCCs development, and that genes found differentially expressed might be candidate biomarkers in these tumors.

In summary, the data presented here have identified genes differentially expressed in HNSCCs from Sudan and Norway. The study has added valuable information by examining HNSCCs from a developing country with population heavily exposed to a known carcinogen, which has been controlled for sampling by established protocols with a developed country. Consistency between the results obtained on Sudan and Norway samples indicates a common biology of HNSCCs, which seems to be mediated by similar pathways regardless of existing differences related to ethnicity, lifestyle, and/or exposure to environmental carcinogens. Our analysis of gene lists reported in a subset of nine similar studies (33–41) indicates that this form of tumor is genetically diverse (with little overlap between gene lists). However, a relatively large

subset of the genes found in at least three of our eight analysis lists were also found in one or more of the other studies. Although many of the genes found have been implicated in HNSCCs, the exact nature of some others warrants further studies. However, and due in part to our choice to use bulk tumors and pooling of samples, additional studies will be needed to address the issue further. Use of bulk tumors, however, as illustrated by this work, does provide substantial information related not only to the tumor itself but also to the molecular events ongoing in the tumor microenvironment. Nevertheless, and due to the large number of genes found and the importance of proper interpretation of global biological significance of detected genes in association with toombak habits and anatomic locations, we plan to test a larger sample size of HNSCCs in an ongoing independent study involving samples of Sudanese patients who are toombak dippers and

cigarette smokers to address the exact roles of these tobacco carcinogens.

Acknowledgments

We thank Inger Ottesen, G. Fjell, and Gunnvor Øijordsbakken for their skilled technical assistance; Bjørn E. Kristiansen, Vidar Steen, Harald Breilid, Aslaug Mugggerud, and Astrid Lgreid (Norwegian Microarray Consortium); Ove Bruland, Benedikte Rosenlund, and Johan Fernø (Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, University of Bergen) for technical assistance; Trond Hellem Bø (Department of Informatics) and Kjell Petersen (Computational Biology Unit, Bergen Centre for Computational Science, University of Bergen, Norway) for their assistance; Prof. Ali M. Idris and personnel (Toombak and Smoking Research Centre, Khartoum, Sudan) and Dr. Abdelaal Saeed (Department of Oral and Maxillofacial Surgery, University of Khartoum, Sudan) for logistic support; and Prof. Saman Warnakulasauriya (Guy's, King's and St Thomas' Dental Institute, King's College London, England) for critically reading and commenting on the article.

References

- Moore SR, Johnson NW, Pierce AM, Wilson DF. The epidemiology of mouth cancer: a review of global incidence [review]. *Oral Dis* 2000a;6:65–74.
- Sankaranarayanan R, Masuyer E, Swaminathan R, Ferlay J, Whelan S. Head and neck cancer: a global perspective on epidemiology and prognosis. *Anticancer Res* 1998;18:4779–86.
- Franceschi S, Bidoli E, Herrero R, Munoz N. Comparison of cancers of the oral cavity and pharynx worldwide: etiological clues. *Oral Oncol* 2000;36:106–15.
- Cancer in Norway 2001. Oslo (Norway): The Cancer Registry of Norway; 2001.
- Ferlay J, Parkin DM, Pissani P, Globacan I. Cancer incidence and mortality worldwide. Lyon: IARC, IARC Press CD Rom; 1998.
- Idris AM, Ahmed HM, Malik MOA. Toombak dipping and cancer of the oral cavity in the Sudan: case-control study. *Int J Cancer* 1995;63:477–80.
- Tralongo V, Rodolico V, Luciani A, Marra G, Daniele E. Prognostic factors in oral squamous cell carcinoma. A review of the literature. *Anticancer Res* 1999;19:3503–10.
- 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 Oncol* 2000;36:256–63.
- Scully C, Field JK, Tanzawa H. Genetic aberrations in oral or head and neck squamous cell carcinoma 2: chromosomal aberrations. *Oral Oncol* 2000;36:311–27.
- Scully C, Field JK, Tanzawa H. Genetic aberrations in oral or head and neck squamous cell carcinoma 3: clinico-pathological applications. *Oral Oncol* 2000;36:404–13.
- Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet (Suppl)* 1999;21:33–7.
- Churchill GA. Fundamentals of experimental design for cDNA microarrays. *Nat Genet (Suppl)* 2002;32:490–5.
- Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. *J Comput Biol* 2000;7:819–37.
- Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002;415:436–42.
- Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
- McDonald MJ, Rosbash M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* 2001;107:567–78.
- Bulyk ML, Huang X, Choo Y, Church GM. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci U S A* 2001;98:7158–63.
- Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999;23:41–6.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
- Pipel Y, Sudarsanam P, Church GM. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* 2001;29:153–9.
- Iyer VR, Eisen MB, Ross DT, et al. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999;283:83–7.
- Slonim DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet (Suppl)* 2002;32:502–8.
- Warner GC, Reis PP, Makitie AA, et al. Current applications of microarrays in head and neck cancer research. *Laryngoscope* 2004;114:241–8.
- Vats A, Tolley NS, Polak JM, Knight BC. Gene expression: a review of clinical applications in otorhinolaryngology-head and neck surgery. *Clin Otolaryngol* 2002;27:291–5.
- Todd R, Wong DT. DNA hybridization arrays for gene expression analysis of human oral cancer. *J Dent Res* 2002;81:89–97.
- Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology in human cancer [review]. *Oncogene* 2003;22:6497–507.
- Cawson RA, Eveson J. Oral pathology and diagnosis. 13. London: Gower; 1987: pp.10–13.
- Dysvik B, Jonassen I. J-Express: exploring gene expression data using Java. *Bioinformatics* 2001;17:369–70.
- Berger JA, Hautaniemi S, Jarvinen AK, Edgren H, Mitra SK, Astola J. Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC Bioinformatics* 2004;5:194.
- Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001;17:520–5.
- Shoop E, Casaes P, Onsongo G, et al. Data exploration tools for the Gene Ontology database. *Bioinformatics* 2004;20:3442–54.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 1999;27:29–34.
- Ginos MA, Page GP, Michalowicz BS, et al. Identification of a gene expression signature associated with recurrent disease in squamous cell carcinoma of the head and neck. *Cancer Res* 2004;64:55–63.
- Belbin TJ, Singh B, Barber I, et al. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. *Cancer Res* 2002;62:1184–90.
- Leethanakul C, Patel V, Gillespie J, et al. Gene expression profiles in squamous cell carcinomas of the oral cavity: use of laser capture microdissection for the construction and analysis of stage-specific cDNA libraries. *Oral Oncol* 2000;36:474–83.
- Al Moustafa AE, Alaoui-Jamali MA, Batist G, et al. Identification of genes associated with head and neck carcinogenesis by cDNA microarray comparison between matched primary normal epithelial and squamous carcinoma cells. *Oncogene* 2002;21:2634–40.
- Roepman P, Wessels LF, Kettelarij N, et al. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. *Nat Genet* 2005;37:182–6.
- O'Donnell RK, Kupferman M, Wei SJ, et al. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene* 2005;24:1244–51.
- Chin D, Boyle GM, Williams RM, et al. Novel markers for poor prognosis in head and neck cancer. *Int J Cancer* 2005;113:789–97.
- Toruner GA, Ulger C, Alkan M, et al. Association between gene expression profile and tumor invasion in oral squamous cell carcinoma. *Cancer Genet Cytogenet* 2004;154:27–35.
- Jeon GA, Lee JS, Patel V, et al. Global gene expression profiles of human head and neck squamous carcinoma cell lines. *Int J Cancer* 2004;112:249–58.
- Diehn M, Sherlock G, Binkley G, et al. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res* 2003;31:219–23.
- Horn EJ, Albor A, Liu Y, et al. RING protein Trim32 associated with skin carcinogenesis has anti-apoptotic and E3-ubiquitin ligase properties. *Carcinogenesis* 2004;25:157–67.
- Choi P, Chen C. Genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma. *Cancer* 2005;104:1113–28.
- Pedri Riento K, Ridley AJ. Rocks: multifunctional kinases in cell behaviour. *Nat Rev Mol Cell Biol* 2003;4:446–56.
- Sorsa T, Tjaderhane L, Salo T. Matrix metalloproteinases (MMPs) in oral diseases. *Oral Dis* 2004;10:311–8.
- Thomas GJ, Jones J, Speight PM. Integrins and oral cancer. *Oral Oncol* 1997;33:381–8.
- Kosmehl H, Berndt A, Strassburger S, et al. Distribution of laminin and fibronectin isoforms in oral mucosa and oral squamous cell carcinoma. *Br J Cancer* 1999;81:1071–9.