Multi-national Gene Expression Profiling of Oral Squamous Cell Carcinomas

Biological Pathways Regardless of Differences Related to Life-style and Ethnicity

Mai Lill Suhr Lunde

Dissertation for the degree philosophiae doctor (PhD) at the University of Bergen

2012

To my family

Contents

Acknowledgements

The work presented in this thesis was performed at the Department of Biomedicine, University of Bergen in the period from 2004 to 2011.

First of all, I would like to thank my supervisors, Associate prof. Salah Osman Ibrahim and Prof. Endre Vasstrand for giving me the opportunity to do this project, for sharing their knowledge and experience and introducing me to the world of oral cancer. My special thanks to Salah for initiating this project and for establishing the international network which made this project possible. Thank you for teaching me so much, your patience, our fruitful discussions and your guidance throughout my studies. Thanks for believing in me!

My special thanks go to Inger Ottesen for her skilled technical assistance, for guiding me in the lab, and for caring so much. Special thanks also to Randi Svebak for being so helpful with the daily lab routines, your knowledge and detailed overview of the lab equipment have been invaluable to me.

I want to thank all my co-authors for making this work possible by supplying oral cancer biopsies, and their contributions to the papers. Special thanks to Prof. Saman Warnakulasuriya for his contributions to all my papers, his interest, comments and ideas, and for reading and commenting on the thesis. Thanks also to Prof. Ravi Mehrotra, Prof. Jan M. Hirsch and Dr. Jussi Laranne for their contributions to my work. Special thanks also to Dr. Ove Bruland for introducing me to quantitative PCR and allowing me to use the facilities at the Centre for Medical Genetics at Haukeland University Hospital. I am also grateful to Dr. Talia Miron for sharing her great expertice with me.

I would like to thank Prof. Ian Pryme for reading and commenting on this thesis.

I could not have done this work without the assistance from the people at the Norwegian Microarray Consortium in Bergen. I would like to thank Harald Breilid and Rita Holdhus for all their help and guidance working with microarray experiments, and Bjarte Dysvik, Kjell Petersen, Anne-Kristin Stavrum, Pawel Sztromwasser and Prof. Inge Jonassen for handling

microarray expression data, patiently explaining the details of data processing and helping me solving problems along the way (there were many!).

I would also like to thank Dana Costea, Gunvor Øijordsbakken and Siren Hammer Østvold for their skilled assistance with immunohistochemistry, and Tarig Osman for taking time to show me the process, and for our interesting discussions.

I highly appreciate working together with my oral cancer group members Dipak Sapkota and Eric Roman, sharing ideas, experience and daily routines. Special thanks to Dipak for helping out with statistics, figures and much more, you are always so helpful! I would also like to thank Bina Raju for sharing social moments and Sushma Bartouli for all our nice lab sessions.

My time at BBB could not be described without mentioning Aurora Martinez and her Biorecognition group, running lab E, and sharing the facilities with our group. Thanks for including and supporting me, without you it would never have been the same! Special thanks to Prof. Aurora Martinez for her generosity, good advice and great support. Thanks so much for giving me the opportunity to participate in one of your projects! My special thanks also to my office mates Jessica and Khanh, making a nice union of Asia, Scandinavia, chemistry, bioinformatics and molecular biology. Thanks to Khanh for always offering me a ride home, and for teaching me so much about Vietnam! And then there is Marte, who has become my dear friend and shared so much with me during these years, both inside and outside of work. Also, thank you for all help in the lab and with figures. You are such a warm and including person, and I am ever grateful to have met you!

To all my old friends here, there and everywhere, thanks so much for being there, caring for me and encouraging me to continue my project. A special thank to Therese and Trude for all your encouragement and support, and Tone for always making my visits home so cheerful! Thanks to Merete and Hanne for providing a social foundation when I first came to Bergen, and to Therese B and the lovely ladies of FFMM for expanding it. I would also like to thank Iren for many pep-talks in the corridors of BBB.

I am very grateful for the support and love of my family, my sister Vivi and brother Ove and their families. I always enjoy our family meetings, and wish they occurred more often. Thanks also for the warmth and care of Heidi and Harald, Kevin and Jill, Wenche and my mother-in-

law Elly, for your kindness and generosity, and my father-in-law Richard and Marit, for making our stays in Lofoten so great. Above all, I thank my parents, Hildur and Erling, for their endless love, encouragement, interest and just being there for me. You have always supported me in all possible ways, and you are wonderful grandparents. I would never have reached this point in life if it weren't for you.

Finally, I want to thank the most important persons in my life, Lars, my dear husband and best friend, for being who you are, for keeping my feet on the ground and for bearing with me, for believing in me and supporting me. You are my safe haven, and a great dad! And to my beloved children, Didrik and Ellinor, you have brought new dimensions into my life, and my love for you has no limits. Life is never boring when you're around!

Thank you all!

Bergen, November 2011

Mai Lill Suhr Lunde

Abbreviations

Array-CGH	Array comparative genomic hybridization
AN	Areca nut
BAC	Bacterial artificial chromosome
BQ	Betel quid
ECM	Extracellular matrix
HME	Hereditary multiple exostoses
HPV	Human papillomavirus
IHC	Immunohistochemistry
MMP	Matrix metalloproteinase
NNN	N-nitrosonornicotine
NNK	4-[methylnitrosoamino]-1-[3-pyridyl] – 1-1 butanone
NPID	Nature pathway interaction database
OC	Oral cancer
OLP	Oral lichen planus
OSF	Oral submucous fibrosis
OSCC	Oral squamous cell carcinoma
PAC	P1 – derived artificial chromosome
PAH	Polycyclic aromatic hydrocarbons
RFLP	Restriction fragment length polymorphism
ROS	Reactive oxygen species
RT-qPCR	Real-time quantitative polymerase chain reaction
SNP	Single nucleotide polymorphism
ST	Smokeless tobacco
TSG	Tumor suppressor gene
TSNAs	Tobacco specific nitrosamines
XME	Xenobiotic-metabolizing enzyme

Summary

Oral squamous cell carcinoma (OSCC) is a major health problem in many developing countries, representing more than 25% of all new cancer cases in some countries such as India and Sri Lanka. Cigarette smoking, smokeless tobacco (ST), alcohol use and chewing of betel quid (BQ) are the main risk factors associated with OSCC development. Infection with high-risk Human papillomavirus (HPV) is an emerging risk factor, particularly for oropharyngeal cancers. Most of the OSCC cases are diagnosed at advanced stages, being one of the factors related to the high mortality rate of this cancer. A better understanding of the molecular biology of OSCC development might lead to improved methods related to detection, assessing prognosis and novel treatments of this malignancy.

Over the recent years, microarray-based technologies have become commonly used techniques for analyzing gene expression and chromosomal alterations in human cancers and other disease conditions. These high-throughput technologies enable genome-wide analysis of changes in gene expression or chromosomal deletions/amplifications in the pathological samples to be studied. In the search for possible molecular biomarkers for OSCCs, a series of studies were carried out in the work described here, where cDNA microarrays and array-comparative genomic hybridization (array-CGH) were applied to examine possible changes in gene expression and DNA copy number alterations in OSCC samples compared to their pair-wised normal controls and between OSCC samples from different populations.

In addition to examining gene expression in OSCC, the study included analysis of samples from a potentially malignant disorder known as oral submucous fibrosis (OSF), common in Asian populations, to ascertain whether genetic aberrations found in OSCC could be observed in early phases of carcinogenesis.

Gene expression profiles and chromosomal alterations were studied in OSCC/OSF samples from Sri Lankan, Indian, Swedish and UK patients using cDNA microarrays and array-CGH. For gene expression profile (**Paper I**), 15 cases of OSCCs from Sri Lanka and their pair-wised normal controls were examined. Following RNA extraction from all samples, cDNA was synthesized and labeled with Cy3 (tumor cDNA) and Cy5 (normal cDNA). Labeled tumor and normal cDNA were hybridized to 31k cDNA microarrays, slides were scanned and images were subjected to analysis with Genepix and J-Express computer software

programmes. 262 genes (189 up-regulated and 73 down-regulated) were found to be differentially expressed between tumors and normal controls with 66 genes of known function and 66 novel genes. Among the group of genes of known function those found were *CAV1*, *CAV2*, *COL4A1*, *MMP1*, *MMP3*, *PLAU*, *SPARC*, *TNC* (all up-regulated) and *AZGP1*, *KRT19* and *S100A1* (all down-regulated). Microarray results for nine genes were verified with RT-qPCR. Hierarchical clustering of the samples based on the differentially expressed genes did not show any clear relationship between sample clustering and the clinicopathological data, except for two samples (one verrucous carcinoma and one advanced tumor) that were clustered separately.

In **Paper II**, gene expression profiles from 19 OSCCs from Sweden (*n=8*) and UK (*n=11*) were examined and compared between these two populations. RNA was extracted from all OSCCs and cDNA was synthesized and labeled with Cy3. For controls, human universal reference RNA was used for cDNA synthesis and labeling with Cy5. Labeled cDNAs were hybridized to 21k human oligonucleotide microarrays, slides were scanned and images were subjected to analysis with GenePix and J-Express. Here, 42 genes (including *APOL3*, *NT5E*, *HMGA1*, *FASN* and *FOS*) were found as being differentially expressed between the two populations compared to controls. Expression of three genes was validated with RT-qPCR. Upon hierarchical clustering, there was a tendency for the samples from the same population to group together.

For chromosomal alterations (**Paper III**) 24 cases of OSCCs (12 from Sri Lanka and 12 from India) and 6 OSF (India) samples were studied. Following total DNA extraction from all samples, tumor and control DNA (Human Universal Reference DNA) were digested and labeled with Cy3 (tumors) and Cy5 (control) and further hybridized to arrays containing 4500 Bacterial artificial chromosomes (BAC) and P1 artificial chromosomes (PAC) clones. Array-CGH resulted in 349 candidate genes located in deleted (34 genes) or amplified (30 genes) different chromosomal regions in these samples common to both populations, in addition to 285 genes located in 66 chromosomal regions found as deleted or amplified in either Indian or Sri Lankan samples. Further, we selected one gene found to be deleted in the samples from both countries, namely *S100A14*, for further validation analysis using immunohistochemistry (IHC) and genetic variation study by RFLP (Restriction fragment length polymorphism). IHC showed decreased expression of S100A14 in OSCC archival samples compared to normal

oral mucosa, and a relocalization from membrane to cytoplasmic expression. RFLP for one SNP (461A>G) demonstrated a significant difference in genotypes of OSCC and OSF.

These results together with our previous findings on cases of OSCCs studied from Sudan and Norway, demonstrate that 72 genes were found to be common for all six populations studied. Among these were *BAX*, *CCND1*, *COL4A1*, *DAPK1*, *FGF3*, *FGF4*, *JUNB*, *MMP1*, *MMP3*, *PLAU*, *SPARC*, *TNC*, *TGFB1*, several S100 gene family members (including *S100A14*) and *TP53*. Of particular note, there were only small differences in gene aberrations in OSF compared to OSCCs.

These results suggest that genetic alterations occur early during OSCC development, and that there are genes commonly involved in OSCC development regardless of the life-style and source of the material to be studied. We suggest *S100A14* as a possible tumor biomarker for OSCCs together with *COL4A1*, *MMP1*, *MMP3*, *PLAU*, *SPARC* and *TNC*. Findings in the present work further suggest that there is a specific genotype of OSF which might be related to an increased risk of OSCC development.

List of publications

Paper I

Gene expression profile of oral squamous cell carcinomas from Sri Lankan betel quid users. Suhr ML, Dysvik B, Bruland O, Warnakulasuriya S, Amaratunga AN, Jonassen I, Vasstrand EN and Ibrahim SO. Oncology Reports 18: 1061-1075 (2007).

Paper II

Gene expression Analysis by cDNA Microarray in Oral Cancers from Two Western Populations. Lunde MS, Warnakulasuriya S, Sand L, Hirsch JM, Vasstrand EN and Ibrahim SO. Anticancer Research, 30: 1083-1092 (2010).

Paper III

Profiling of chromosomal changes in potentially malignant and malignant oral mucosal lesions from South and South-East Asia using Array - Comparative Genomic Hybridization. Lunde MS, Roman E, Warnakulasuriya S, Mehrotra R, Laranne J, Vasstrand EN and Ibrahim SO (2011). Manuscript (Submitted to BMC Cancer).

1. Introduction

1.1 Oral Cancer

Oral cancer (OC), accounting for about 40% of all head and neck cancers, is usually defined as a malignancy affecting the lip, tongue and oral cavity, floor of mouth, gingivae, buccal mucosa, oral commissures, retromolar trigone and hard palate (1, 2). More than 90% of all OC cases are OSCCs, representing a homogenous group of neoplasms with common risk factors and similarities in clinical presentation, prognosis and treatment (2, 3). OSCC develops through a series of cellular changes, initiated by single cell epithelial alterations (atypia), followed by dysplasia, involving more cells, progressing into carcinoma *in situ* and finally invasion. Epithelial alterations and development of dysplastic lesions are associated with OSCC, as these disorders are likely to progress to carcinoma. Leukoplakia, erythroplakia, OSF and oral lichen planus (OLP) are among the common oral mucosal disorders which have been described under the terms "pre-malignant", "pre-cancers", "precursor lesions" and intraepithelial neoplasia". Recently, the WHO has recommended that the term "potentially malignant oral lesions" is to be used to describe these conditions (4).

Worldwide, the incidence of OSCC varies greatly, with a relative frequency extending from 1-2% in developed countries like United Kingdom and Sweden, to more than 30% in developing countries like Sri Lanka and India. In South-East Asia, OSCC constitutes a major health problem, being the fifth most common form of cancer (2, 5). Nevertheless, the incidence of OSCC is increasing also in the developed world, particularly among younger people (2). Mortality of OSCC also has geographical variations, but overall the five-year survival rate is less than 50%, and has remained quite stable for the last four decades.

The large variations in geographical incidence of OSCC are influenced by etiological risk factors such as the use of smoked tobacco and ST and alcohol, the main causative factors for this disease. In the Indian subcontinent, the use of BQ chewing and bidi smoking is extensive (2, 6), along with alcohol drinking (7). BQ consists of a betel leaf packed with sliced areca nut (AN), tobacco (not always), slaked lime and sometimes sweeteners (8, 9), while bidi is a cigarette rolled in a dried temburini leaf. BQ chewing is common among men, women and children, accounting for a major part of the clinical cases of potentially malignant disorders and OSCC in this region (7). Cigarette smoking and alcohol consumption are major risk factors for OSCC in Western Europe. In Sweden, the habit of oral snuff ("*snus*") use is

common. Although recent studies have linked snuff use to development of OSCC, others have excluded this habit as an OSCC risk factor (10-15). However, tobacco and alcohol may act synergistically in OSCC development, increasing the risk significantly, as much as 38 times for heavy drinkers and smokers compared to non smokers and drinkers (2). Other risk factors associated with OSCC development are nutritional deficiencies, human papillomavirus (HPV) infection (2, 16) and syphilis. Some dental factors such as chronic trauma, periodontal disease and poor oral hygiene are also factors implicated in a few studies (17).

It is clear that there are world-wide differences related to incidence and etiology of OSCCs. In table 1, suggested incidence rate(s) and etiological factors related to OSCC development in UK, Sweden, India and Sri Lanka are presented.

Table 1. Etiology and relative frequency of OSCC in UK, Sweden, India and Sri Lanka

Etiology/Epidemiology	**UK**	**Sweden**	**India**	**Sri Lanka**
Relative frequency (%)	1-2	1-2	>30	>30
Major risk factors	Smoking	Smoking	BQ	BQ
	Alcohol	Alcohol	Smoking	Smoking
	Socio-economical status	Snuff?	Alcohol	Alcohol
			Poor nutrition	Poor nutrition
	HPV?		Insufficient health care	Insufficient health care

1.2 Potentially malignant oral mucosal lesions

1.2.1 Oral submucous fibrosis

OSCC progresses through a multistep process initialized by genetic alterations, often arising from the malignant transformation of a premalignant oral mucosal condition such as OSF, initially described as an insidious, precancerous chronic disease that may affect the entire oral cavity, sometimes extending to the pharynx (18). OSF is characterized by inflammation in the subepithelial tissue, followed by fibrosis of the submucous tissues, resulting in stiffness of the oral mucosa and difficulties in opening the mouth (Figure 1a).

AN and BQ chewing are the most common risk factors associated with OSF, and the condition is predominantly seen in the South Asian regions where use of AN and BQ are common. Immigrants from South Asian countries living in western communities and sustaining these habits may also suffer from OSF (8, 19).

BQ users who develop OSF may have functional difficulties in eating, swallowing and speech, and experience a burning sensation of the oral mucosa. Palpable fibrous bands, a marble-like appearance, blanching of the mucosa and mucosal ulceration are characteristic features used in diagnosing OSF. This pathological condition is not reversible, unless cessation of AN/BQ chewing is achieved at a very early stage. OSF has a high potential for malignant transformation where an Indian cohort study reported a malignant transformation rate of 7.6% over a period of 17 years (20).

1.2.2 Oral leukoplakia

Oral leukoplakia is a potentially malignant lesion that presents clinically in the form of a white patch (Figure 1b). WHO describes leukoplakia as "white plaques of questionable risk having excluded (other) known diseases or disorders that carry no increased risk of oral cancer." Leukoplakia is only a clinical term, and conditions described as leukoplakia can be given another diagnosis after biopsy analysis (21, 22). Leukoplakias are seen most frequently on the buccal mucosa, alveolar mucosa and lower lip, and men are more often affected than women (increasing with age). Dysplastic or malignant transformations occur more often in lesions localized in the lower lip, lateral border of the tongue and floor of the mouth (21). There are two types of oral leukoplakias, non-homogenous and homogenous. Thin leukoplakias with an even surface are homogenous, while thicker and verrucous lesions are non-homogenous. Mixed white and red lesions, called erythroleukoplakias, have an uneven surface and are also non-homogenous. Generally, the non-homogenous leukoplakias are more likely to undergo malignant transformation (22). There are global variations in the prevalence of oral leukoplakia, but a mean global prevalence of 2.6% has been reported (23). The frequency of leukoplakia undergoing malignant transformation is not consistent in the studies performed, but in relation to a prevalence of 2.6%, the global transformation rate has been estimated to be 1.36% (23).

1.2.3 Oral erythroplakia

Erythroplakias are defined in the same way as leukoplakias, as a red patch that cannot be defined clinically or pathologically as any other condition (Figure 1c). Erythroplakias are seen more often in older men, and the most common sites are floor of the mouth, lateral border of the tongue, retromolar pad and soft palate (21, 22). Erythroplakias are less common compared to leukoplakias, but are more likely to progress to malignant lesions. Sometimes

erythroleukoplakias are seen, the red lesions inter-mixed with white lesions, and the probability that the red parts will show a higher grade of dysplasia is higher than for the white component (22).

1.2.4 Oral lichen planus

Oral lichen planus (OLP) is a chronic inflammatory condition of unknown etiology, affecting the oral mucous membranes (Figure 1d) (24, 25). OLP can also affect other sites of stratified squamous epithelia like skin and genitalia. It is a T cell-mediated autoimmune disease, perhaps associated with stress but of unknown etiology. Increased Th1 cytokine production in combination with activated T cells increase the expression of intercellular adhesion molecules, attracting the T cells towards the oral epithelium (24, 25). Up-regulation of the epithelial basement membrane extracellular matrix proteins further attracts T cells. The T cells bind to keratinocytes and interferon-gamma (*INF-ɣ*), followed by up-regulation of p53 and matrix metalloproteinases (MMPs), leading to destruction (apoptosis) of the basal keratinocytes (24). Genetic polymorphisms of INF-γ and tumor necrosis factor-alpha (*TNF-α*) are associated with increased risk of OLP, and localization of lesions (25, 26).

Morphologically, OLP is classified as erosive or reticular (keratotic), often with multiple morphologies presented simultaneously (25). OLP are chronic disorders with no definite cure. A significant risk of malignant transformation of OLPs has been demonstrated, and patients with OLP should be monitored on a long term basis (25, 26) .

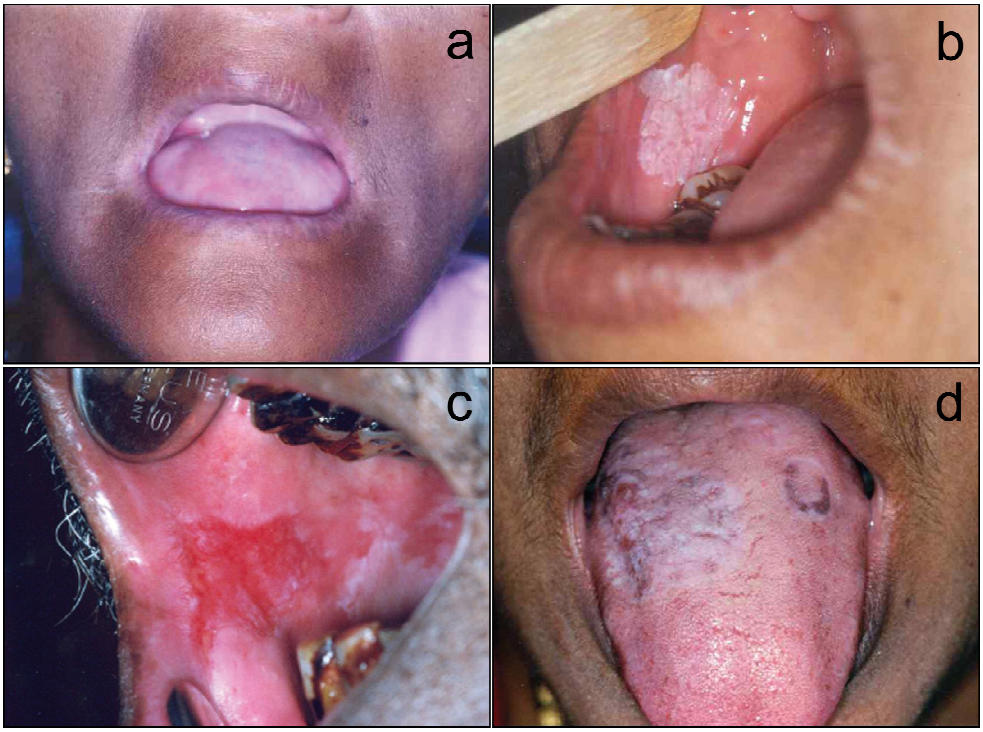

Figure 1. Potentially malignant oral mucosal lesions. a) Oral submucous fibrosis. The patient has restricted mouth opening, and blanching of the tongue; b) Homogenous leukoplakia; c) Erythroplakia. d) Lichen planus. Two patches, one on the right side of the dorsum of the tongue intermingled with areas of pigmentation, and another smaller one on the left side. IARC (htpp://screening.iarc.fr)

1.2 Tobacco habits and development of OSCCs

Tobacco use in the form of ST and smoked tobacco is a major risk factor for human cancer development, being the single largest cause of death due to preventable cancer (27). Cigarette smoking is the cause of 90% of all lung cancers, the most common cancer in the world (27). For OSCC development, tobacco use is the main risk factor world-wide. In the Western World, cigarette smoking is predominant, while tobacco chewing and oral snuff use are the common tobacco habits in developing regions (28, 29).

Cigarette smoke contains more than 60 known carcinogens, both in mainstream and sidestream smoke (27). Some of the strong carcinogens include polycyclic aromatic hydrocarbons (PAHs) like benzo[a]pyrene, tobacco specific nitrosamines (TNSAs) like N-nitrosonornicotine (NNN) and 4-[methylnitrosoamino]-1-[3-pyridyl]-1-1 butanonone (NNK) and aromatic amines 4-aminobiphenyl (27, 29, 30). These compounds are present in smaller amounts than the weaker carcinogens like aldehydes (formaldehyde, acetaldehyde) and

phenols (cathecol) (27). The most important mechanism for mutagenesis caused by tobacco carcinogens is DNA binding, resulting in DNA adducts and miscoding if not repaired (27).

ST products are very heterogeneous in the way they are produced, marketed, their content and how they are consumed (30). Some products are chewed, while others are used as snuff. Chewing tobacco comes in the form of loose-leaf tobacco, cut or shredded, and treated with flavoring solutions. This product is declining in popularity, while the use of moist snuff is increasing in developing countries (28, 31). Moist snuff consists of finely cut tobacco, either loose or packed in small portion bags. The latter is user-friendly, and has become very popular in the US over the last 15 years (30). Snuff may also be used nasally in the form of pulverized dry snuff, but the use is not widespread. ST does not contain as many carcinogens as cigarette smoke, due to the fact that carcinogens are formed during combustion. However, ST has a considerably higher content of TSNAs than tobacco smoke, and levels of aldehydes and metals are significant (27, 30). In the European Union, oral ST products are banned, with the exception of Sweden, where oral snuff is very popular (30, 31).

Because of the manufacturing methods of Swedish snuff, resulting in a virtually sterile product, it contains a lower amount of bacterial by-products like TSNAs than North American moist snuff. Although evidence for Swedish snuff to be considered as a risk factor for OSCC is available, some studies have shown the opposite (10, 12, 14, 32, 33). Snuff is even suggested as a replacement for cigarette users addicted to nicotine (14, 27, 28). There is, however, a significant risk of developing pancreatic cancer in association with snuff use (32, 34).

In the Sudan, oral snuff use, locally called *toombak*, is also widespread, and has been used for more than 400 years (35). *Toombak* consists of tobacco powder of *Nicotiana rustica*, mixed with aqueous sodium bicarbonate solution, producing a loose moist snuff dipped and retained between mostly the gums and the lower lips, and less often in cheeks or floor of the mouth (36). It is sucked slowly for 10-15 minutes and replaced several times per day (36). This type of snuff is different from the Swedish and American snuff regarding manufacturing procedures, tobacco species and nitrosamine content (36, 37). The level of TSNAs present in *toombak* are much higher than in Swedish and American snuff (37, 38), and *toombak* is considered to play a major role in the etiology of OSCC with the tumors often localized to the site of dipping in the mouth (39). Although *toombak* is used all over the Sudan, the habit is much more common among men rather than women, and therefore a higher rate of OSCCs is seen in Sudanese men (35, 39).

In India, Sri Lanka and other South East Asian countries the use of BQ (also called paan) is extensive, and has been for centuries (8, 9, 40). BQ consists of sliced AN from the Areca palm (*Areca cathecu*), cathecu (*Acacia catechu*) and slaked lime wrapped in a betel leaf (*Piper betle*). Often sweeteners and spices are added to improve the taste (8, 9, 40). In India and Sri Lanka, adding tobacco to the BQ is common, while this is not done in Taiwan, Papua New Guinea and China (6, 41). Regardless of the regional variations in preparing BQ, AN is the most important ingredient. There are commercial BQ variants called pan masala and guthka, heavily marketed contributing to an increase in BQ use (8). Some BQ products are made especially for children, called sweet supari, gua or miste paan (19). The quid is placed between the teeth and the buccal mucosa, and sucked or chewed for several hours, producing a feeling of euphoria and well-being (19).

Being an ancient habit, BQ chewing is socially accepted both among men, women and children. On a daily basis, BQ is used by millions of people, representing a major risk factor for developing OSCC. The incidence of OSCC is very high in regions where BQ chewing is common, and there is undoubtedly an association between BQ and OSCC development (6, 40). Also, it is a known risk factor for OSF (6, 8, 9).

Carcinogens present in BQ are the tobacco-specific nitrosamines NNN and NNK as in other tobacco products (8, 37). Also, AN contains nitrosamines, *N*-nitroso-guvacoline (NG) and 3-(methyl-*N*-nitrosamino)propionitrile (MNPN), the latter being carcinogenic (8, 9). Alkaloids such as arecoline, arecaidine, guvacoline and guvasine have also been reported as areca nut carcinogens. BQ chewing generates reactive oxygen species (ROS) like superoxide anion and hydrogen peroxide, due to the combination of slaked lime and areca nut. The lime (calcium hydroxide) increases the pH in the oral cavity to alkaline conditions, favoring ROS formation, causing oxidative damage to the epithelial cells (8, 9). Decreased lime content of BQ should reduce ROS formation (8).

Being such a popular habit among both young and older people, linked to OSF and OSCC development, awareness of the danger of this habit should be emphasized, and preventative actions should be taken.

1. 4 Molecular development of OSCCs

Cancer is a genetic disease generally initiated by a mutation in a single cell, although the development of an invasive cancer is a result of accumulation of further genetic alterations

(42, 43). Usually, the tumor develops during a long period of time, where the cancer cells arise from generation of cells progressing into tumorigenesis (42, 43). Cancer cells acquire new traits during this evolutionary process, making them able to avoid the anticancer defense mechanisms harbored by normal cells (44). These traits have been described as the "Hallmarks of Cancer", and include self-suffiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, immortalization, sustained angiogenesis, and invasion and metastasis (44). There are certain germ line mutations responsible for increased susceptibility to OSCC, but most of the mutations in the tumor cells are somatic (42, 43). Carcinogens, especially from tobacco and alcohol use, can in the case of OSCC development cause mutations in proto-oncogenes, tumor suppressor genes (TSGs) or DNA repair genes, leading to deregulation of cell growth, inability to repair damaged DNA and/or to eliminate damaged cells (27, 30, 31). Important tobacco carcinogens are TSNAs like NNN and NNK, which particularly cause guanidine-to-thymidine transversion (27). PAHs are tobacco related, but also an environmental carcinogen (27, 45). Ethanol per se is not carcinogenic, but acetaldehyde arising from alcohol metabolism is a recognized carcinogen (45).

There are large variations in the individual ability to metabolise carcinogens, partly due to the polymorphic nature of xenobiotic-metabolising enxymes (XMEs), and bearers of certain polymorphisms may have a predisposition to OSCC (45). Glutathione-S-transferases (GSTs) are a group of XMEs involved in detoxification of carcinogenic compounds, but overexpression of GSTs may also lead to increased resistance to anti-cancer drugs (46). There are several studies relating different polymorphisms of GST to OSCC susceptibility and predisposition to oral malignant disorders (45, 47-49).

Genetic alterations can arise from point mutations in a single gene, chromosomal translocations, deletions and amplifications of chromosomal regions and epigenetic silencing such as methylation and transcriptional repression (43). Mutations in the proto-oncogenes (like *RAS*, *MYC*, *CCND1,* and *EGFR*) will result in activation of genes stimulating cell growth and avoiding apoptosis (gain of function). In normal cells, these genes are tightly regulated to control cell proliferation, but when a mutation arises, an oncogene (cancer-causing gene) is activated, and the cells are allowed to proliferate out of control (43, 50). Point mutations, chromosomal rearrangements and amplifications are common mechanisms activating oncogenes. Tumor suppressor genes (TSGs) (*RB*, *TP53*, $p16^{INK4A}$/$p14^{ARF}$ and *PTEN*) have reduced or lost activity (loss of function) through mechanism(s) like loss of heterozygosity (LOH), epigenetic events, mutations or deletions (43, 50). Silencing of the TSGs leads to loss of control mechanisms of cell growth arrest and apoptosis, preventing repair of DNA damage

and elimination of abnormal cells, leading to accumulation of genetic alterations and eventually tumor formation (43, 50). Genes that are responsible for DNA damage repair and elimination of damaged cells (stability genes) are represented by *ATM*, *BRCA1*, *BRCA2*, *BLM* and *NBS1*. The stability genes are involved in control of processes such as DNA damage repair, chromosomal segregation and mitotic recombination, resulting in a higher mutation rate when inactivated (50).

In OSCCs, alterations in more than 100 genes have been described, but mutations in proto-oncogenes and TSGs are found to be of great importance in oral tumorigenesis (16, 51, 52). Figure 2 shows a suggested progression model for OSCC. An early event of OSCC development is LOH on chromosome 9p21-22, seen in 70 to 80% of tumors (51, 53). This region harbors *CDKN2A*, a TSG encoding two different proteins, p16INK4A and p14ARF. These proteins regulate cell cycle G1/S progression and stabilize p53 by MDM2 binding (16, 54). Chromosome 3p, 13q, 17p and 11q are also involved in early tumorigenesis, including alterations in 17p13.1, where *TP53* gene is located. p53 is one of the proteins most frequently inactivated in many human cancers, including OSCCs (51, 55, 56). *TP53* mutations seem to be more frequent in OSCCs from Western countries than in developing countries (7, 57, 58), perhaps related to the different amounts of carcinogens found in cigarette smoke and ST (27, 58). In oral dysplastic lesions, *TP*53 mutations are also found to be common, demonstrating early involvement of p53 in OSCC development (16, 58, 59). p53 may also be inactivated through other mechanisms such as degradation by MDM2, MDM4 or other p53 regulators, or by viral proteins (51, 60). Oral human papillomavirus (HPV) infection is considered as a risk factor for OSCC development, as its presence may lead to inactivation of p53. There is also a possible relation between HPV infections in OSCC and *HRAS* oncogene mutation, located in chromosome 11q (51). *HRAS* mutations are seen mainly in the developing world, with a particularly high frequency in India (51). Cyclin D1 is another 11q oncogene, often amplified in oral premalignant mucosal lesions and OSCCs, affecting cell cycle G1/S transition (51, 59). The first tumor suppressor gene to be identified, the retinoblastoma gene, *RB*, is located in chromosome 13q, and has shown to be frequently lost in OSCCs and also in oral dysplastic lesions (51, 59, 61). These genetic changes are connected with early events in OSCC tumorigenesis, leading to the progression of dysplasia into carcinoma *in situ* (51, 53, 59, 61). Further progression into invasive carcinoma involves genes like *MMP*s, genes related to cell adhesion, proliferation and migration and *TP53* gene mutations which also occur in this more advanced stage (51, 53). Overall, chromosomal deletions at 1q, 3q, 1p, 3p, 4p, 5p, 7q, 8p, 10p, 11q, 13q, and 18q, with amplification at chromosomes 1q, 3q, 5p, 7q, 8q, 9q, 11q, 12p, 14q

and 15q are the common genetic alterations found in OSCC (16, 54). Alterations accumulating over a time period of 20 to 30 years will eventually lead to the development of OSCC (62).

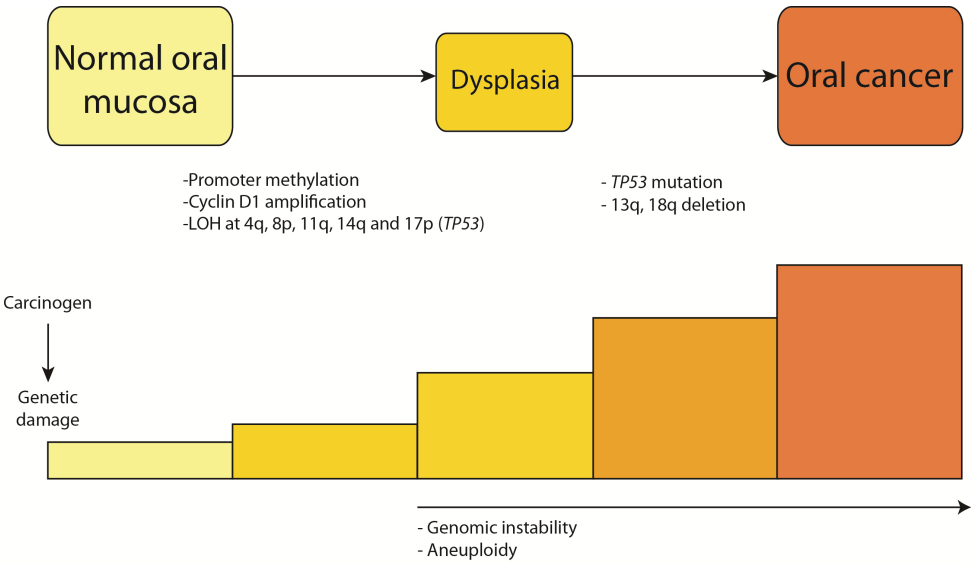

Figure 2. Suggested progression model for OSCC development. Adapted from Califano *et al.* (53) and Choi *et al.* (59).

1. 5 The S100 proteins

The S100 proteins are small acidic proteins (molecular mass 10-20 kDa) of the EF-hand superfamily, being involved in a large number of cellular processes like cell cycle regulation, cell growth, cell differentiation and cell motility (63, 64). The S100 proteins are expressed both intracellularly and extracellularly (63, 64). There are at least 25 members in the S100 protein family, and several members (A1-A18, among others) are clustered in chromosome 1q21, a region often affected by alterations in human cancers (65, 66). The S100 proteins are exclusive for vertebrates, first identified in bovine brain by Moore in 1965, and named S100 because the subcellular fraction was *S*oluble in 100% saturated ammonium sulphate (67). The nomenclature of S100 genes is complex because many of the members have several names in the literature (63).

EF-hand motifs are composed of two helices, E and F, joined by a loop containing the site for Ca^{2+} binding (Figure 3). S100 proteins are symmetric dimers containing two EF-hands motifs connected by a central hinge region (68, 69). The C-terminal EF-hand contains the typical EF-hand calcium-binding motif, common to all EF hand proteins. The Ca^{2+}-binding motif has a 12 amino acid sequence, flanked by helices III and IV. The N-terminal EF-hand is characteristic of the S100 proteins (called pseudo-EF-hand) and consists of a specific 14 amino acid sequence, flanked by helices I and II. A stretch of amino acids is located subsequent to the C-terminal end, the C-terminal extension. The hinge region and the C-terminal extension are the most variable regions of S100 proteins, and are linked to their different biological properties (63, 68, 69). When Ca^{2+} binds to the S100 proteins, a conformational change occurs, mostly in the C-terminal region. A hydrophobic surface is formed by the hinge region, helix III and the C-terminal loop region, allowing target proteins to interact with the S100 protein. In addition to calcium ion binding, many S100 proteins have high affinity for Zn^{2+}, and Cu^{2+} (63, 68, 69).

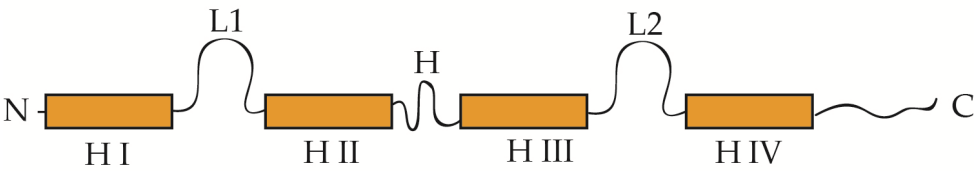

Figure 3. Schematic structure of the S100 proteins. HI-IV: Hinge regions; H: Central hinge region; L: Loop. Adapted from Donato (68).

The S100 proteins are signaling molecules with a broad functional diversity. Intracellular functions of S100 proteins are regulation of phosphorylation, enzyme activity, cytoskeleton organization, Ca^{2+} homeostasis and cell growth and differentiation (63, 70). Extracellular functions are of a cytokine-like manner, acting as chemoattractants and exhibiting neurotrophic activity. Through interacting with the receptor for advanced glycation end products (RAGE), S100 proteins activate intracellular signaling pathways both promoting cell survival and apoptosis (63, 68, 71). S100 proteins are also involved in metastasis and cell proliferation. S100A4 has been linked to a number of malignancies such as thyroid carcinoma, melanoma, breast carcinoma, lung cancer among others, playing a significant role in metastasis and tumorigenesis (64). Other S100 genes implicated in metastasis are *S100A7*, *S100A8* and *S100A9*. Some S100 proteins are over-expressed in cancerous tissue, while others

are down-regulated. Some S100 genes act as tumor suppressors, while others are tumor promoting, even the same gene can have different roles in different cancer forms (63).

1.5.1 S100 proteins and OSCCs

Several of the S100 proteins are linked to different human cancers, both in cancer progression and prevention (63, 64, 71). There are a number of S100 proteins interacting with p53, like S100A4 and S100B, both inhibiting p53 activity possibly through inhibition of p53 phosphorylation. S100A4 has also been demonstrated to enhance p53-dependent apoptosis (72, 73). S100A2 also binds to p53, and has been demonstrated to promote p53 transcriptional activity (74, 75), and in our recent work we demonstrated that over-expressing *S100A14* in OSCC-derived cell lines resulted in nuclear accumulation of p53, suggesting that S100A14 protein enhances stabilization of p53 activity (76). In addition, Chen *et al.* demonstrated that *S100A14* is a downstream gene of *TP53*, down-regulated in esophageal cancer and a possible tumor suppressor in this malignancy (77). *S100A14* was found to be down-regulated in OSCCs in our earlier work (78) and in another study on esophageal cancer (79). Apart from *S100A14*, other S100 genes have been shown to be implicated in OSCC. Our previous work demonstrated down-regulated expression of *S100A2*, *S100A4*, *S100A6*, *S100A8* (and *S100A14*) in OSCCs compared to normal tissue (78, 80), and amplification of the chromosomal region 1q21 harboring *S100A1- 16* (81). It has been suggested that S100A2 may act as a tumor suppressor in OSCC (82, 83). Down-regulation of S100A6 has been reported in OSCC and OSCCs-derived cell lines (84). *S100A7* have been found as up-regulated in premaligant oral mucosal lesions and early-stage oral carcinomas (85-88), proposing a role for S100A7 in malignant transformation of premalignant conditions. An association between S100A8 and HPV18-infected OSCCs was demonstrated by Lo *et al.* (89), suggesting S100A8 as a biomarker in HVP-associated OSCCs. S100A8, S100A9, S100A11 and S100A13 proteins were found to be differentially expressed between OSCCs and normal controls by MALDI-TOF MS analysis (90).

The *S100A14* gene has recently been demonstrated as a possible tumor suppressor by Chen *et al.* and our group (76, 77). The protein was first identified from a lung cancer cell line by Pietas *et al.* in 2002 (91), as a protein with molecular weight of 11,66 kDa containing two helix-loop-helix structural motifs characteristic of calcium-binding sites. S100A14 protein shares 68% similarity and 38% identity with S100A13, 62% similarity and 30% identity with S100A4, 58% similarity and 31% identity with S100A10 and 55% similarity and 34% with S100A9. The gene is located on chromosome 1q21 and consists of three introns and four

exons. S100A14 is expressed in a number of normal human tissues, like colon, breast, liver, ovary and prostate. In cancerous tissue, S100A14 has been reported as over-expressed in tumors such as ovarian, breast, uterus and prostate (91), but down-regulated in esophageal, kidney, rectum and colon tumors. In OSCCs, S100A14 was demonstrated as down-regulated, and is suggested to be involved in tumorigenesis. An interaction between S100A14 and p53 protein was shown in esophageal cancer (77), and in our previous work, where we found that S100A14 is involved in regulation of p53 and cell cycle in OSCC-derived cell lines (76). Also, we demonstrated that S100A14 is involved in invasion and metastasis of OSCC-derived cell lines (92). To date, the knowledge of the biological functions of the S100A14 protein is limited, and closer studies are needed to reveal its role in normal tissue and in oral tumorigenesis.

1. 6 Genetic changes and chromosomal damages in OSCCs: Role of array technology

During the recent years microarrays have become powerful and widely used tools to study the expression of thousands of genes simultaneously. Microarrays are platforms of different materials (glass, silica, nylon, beads etc) containing oligonucleotides or cDNA probes that are spotted or synthesized onto the platform surface (93-96). There are numerous commercial platforms available today, for several organisms, and custom-made arrays are also available (96).

The general purpose of applying microarray technology is to search for a change in gene expression (or DNA copy numbers) between different groups of samples/patients, like diseased compared to healthy, treated compared to non-treated and pre-cancer compared to cancer. Also, rather than searching for changes in single genes, microarrays are used to determine specific genetic profiles or expression patterns which may be linked to certain conditions such as tumor classification, tumor aggressiveness or response to treatment (95, 97).

There are both one- and two-channel microarrays available. The two-channel systems were the first arrays to be used, but today the one-channel systems are used more frequently. Figure 4 illustrates the two-channel system, where cDNA (synthesized from RNA) from the test samples and the control samples are labeled with two different fluorescent dyes, and hybridized to a microarray slide. cDNA will bind to the individual spots on the microarray in

different levels according to the amount of mRNA present in the samples. The amounts of mRNA present are linked to the signal intensity of each spot, and the signal intensity is a measurement of the expression of the specific gene in the test sample compared to the control sample. Upon scanning the arrays, the data need to be filtered and normalized before they are further analyzed.

A literature search on microarrays and gene expression profile of OSCCs showed numerous results and findings with different aspects of changes related to OSCC gene expression profiles. Differential gene expression between tumors and normal controls have been studied to a large extent, and this has shown that many common genes and pathways are found to be associated with OSCC development (97, 98), including our earlier studies (80, 99). Several studies focusing on different stages of OSCC development have provided important information about tumor classification and of oral premalignant conditions compared to OSCCs (100-105) or metastasizing compared to non-metastasizing tumors (106-113). A study on differential gene expression between OSCCs from smokers and BQ chewers have also been performed (114). The findings of molecular biomarkers and/or genetic profiles associated with malignant transformation of potentially malignant conditions and prediction of metastatic potential are very important for the understanding of molecular events involved in OSCC development, and may lead to improved detection and treatment strategies of the disease.

Array- based comparative genomic hybridization (array-CGH) enables genome-wide analysis of chromosomal deletions and amplifications in one single experiment, at a much higher resolution than conventional metaphase spread CGH, which detects chromosomal alterations of relatively large regions (~10Mb). Development of array-based CGH has opened for the detection of chromosomal rearrangements at levels less than 1Mb (94, 115, 116). Array-CGH follows the same principle as the described procedure for microarray experiments, but the samples used for labeling and hybridization are extracted DNA. The probes on the arrays consists of either large insert clones such as BACs, cDNA or oligonucleotide sequences (94, 116, 117).

DNA copy number variations are associated with many pathological conditions, and array-CGH has become a valuable tool to detect chromosomal alterations. Clinically, the technique has been used to study abnormalities like autism, schizophrenia and other mental disorders, developmental delay, dysmorphic features and other syndromic conditions (115, 117, 118). This technique is also widely used to study copy number variations in cancer (94, 115, 118, 119). A study on the global variations of DNA copy number concluded that there

are large variations between African, Asian and European populations (120). In OSCC and head and neck SCC, conventional and array-CGH has been applied to search for chromosomal alterations and possible biomarkers in tumors compared to normal tissues showing chromosomal changes similar for many tumors (81, 121-126). Studies have also been performed on classification of tumor type, tumor stages and prognosis with findings of specific copy number variations for different types of tumors and oral premalignant disorders compared to OSCCs (127-129). One study demonstrated copy number variations in premaligant lesions associated with a high risk of malignant transformation (130). Two studies performed with BQ-related OSCCs showed chromosomal changes that might be related to BQ (131, 132).

As both microarrays and array-based CGH technologies have been developed and improved, new findings in OSCC tumorigenesis have been brought about, and microarrays are established research tools both for cancer, diseases in general and other research areas.

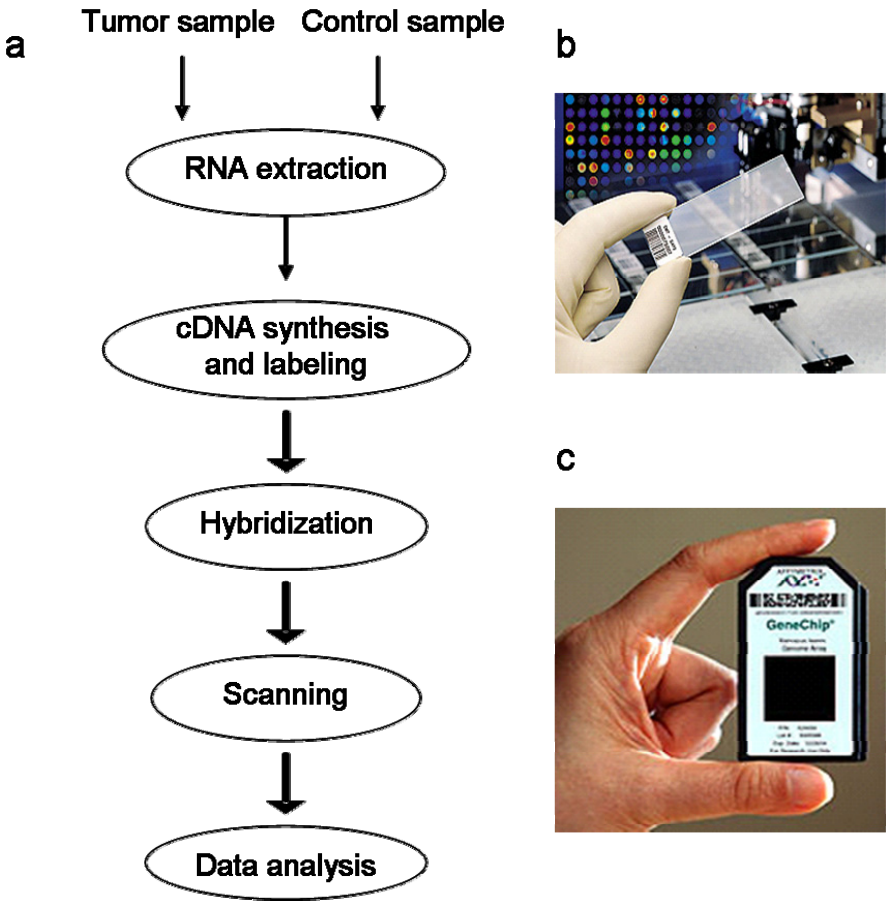

Figure 4. Overview of the microarray process. a) Flowchart of the microarray procedure. b) A glass microarray, probes are spotted onto the surface. c) Affymetrix genechip, probes are synthesized onto the surface of the array.

1. 7 Real-Time quantitative polymerase chain reaction (RT-Q-PCR)

The real-time quantitative PCR technique is a well established method for measuring gene expression and validation of microarray experiments. RT-qPCR enables detection and quantification of the target molecule since the progression of the PCR reaction is monitored as it occurs in real-time. With traditional PCR, the product is detected at the end of the reaction, *i.e.* in the plateau phase (Figure 5). At this point, the PCR products will have differences in quantity due to differences in reaction kinetics. In RT-PCR, the detection of the PCR products occurs in the exponential phase, where the PCR products are amplified exponentially (doubled) in every cycle. This ensures a precise quantification of the PCR products. The point

at which the fluorescence in a reaction reaches a level above the background is called the cycle or crossing threshold Ct (Figure 6) (133, 134).

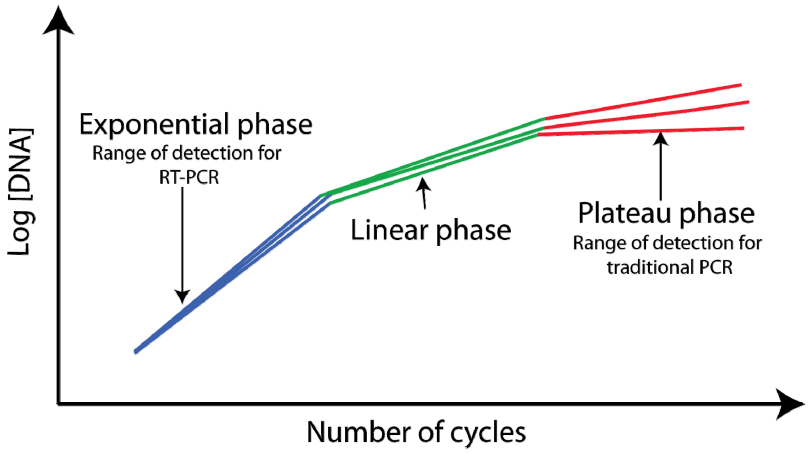

Figure 5. The three phases of PCR, showing three replicates of the same sample. In the exponential phase, where the products are detected in RT-PCR, the products are doubled with each cycle, when the reagents are fresh and available. In the linear phase, some of the reagents are being consumed, and three samples begin to diverge in their quantities. In the plateau phase, the reactions will end, at different times for each sample due to reaction kinetics. Adapted from Applied Biosystems.

RT-qPCR is frequently combined with reverse transcription, when the quantification target molecules are RNA. The products of RT-PCR are detected by using fluorescent probes, and the reaction is performed in a thermocycler. There are several probes available based on different detection techniques, such as hydrolysation of the probe (Taqman), hybridization (Lightcycler), fluorescent hairpins (LUX) or intercalating probes (SYBR Green). The Taqman assay will be described more closely, as this is the technique that has been used in our studies.

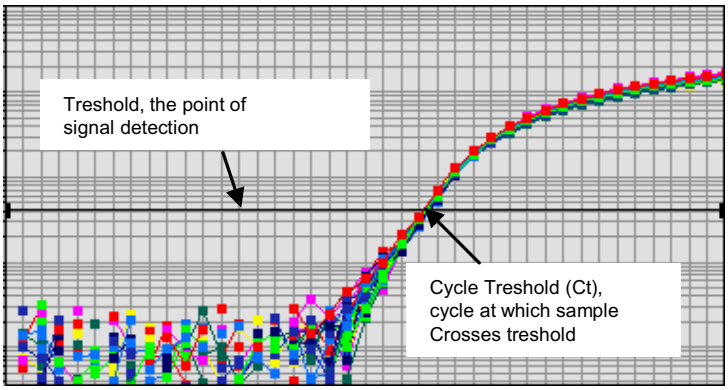

Figure 6. Amplification curve, showing the number of cycles on the horizontal line and fluorescence signal on the vertical line. (Applied Biosystems)

The Taqman assay is based on an oligonucleotide probe containing a reporter fluorescent dye on the 5'-end and a quencher dye on the 3'-end. When the two dyes are close on the probe, fluorescence resonance energy transfer (FRET) through space will limit the signal from the reporter dye. The probe anneals to the target sequence, and is cleaved by the 5' nuclease activity of Taq DNA polymerase as this primer is extended, separating the reporter and quencher dyes, increasing the reporter dye signal (and decreasing the quencher dye signal) (Figure 7). The reporter signal will increase proportionally with the amount of PCR product made in each cycle.

The RT-qPCR results can be analyzed either by absolute quantification, the absolute number of a specific RNA per sample, or by relative quantification, assessing the relative difference of RNA copies between samples. Absolute quantification requires a standard curve of known amount of RNA (using a serial dilution). The unknown signals are quantified by interpolating their quantity from the standard curve. A digital method not requiring a standard curve is also available (133, 134). The standard curve may also be applied in relative quantification, but the most common method is the comparative Ct method, using the formula $2^{-\Delta\Delta CT}$. This method compares the Ct value of the gene of interest to an endogenous control. The efficiency of the target gene amplification must be approximately equal to the efficiency of the endogenous gene amplification (133, 134).

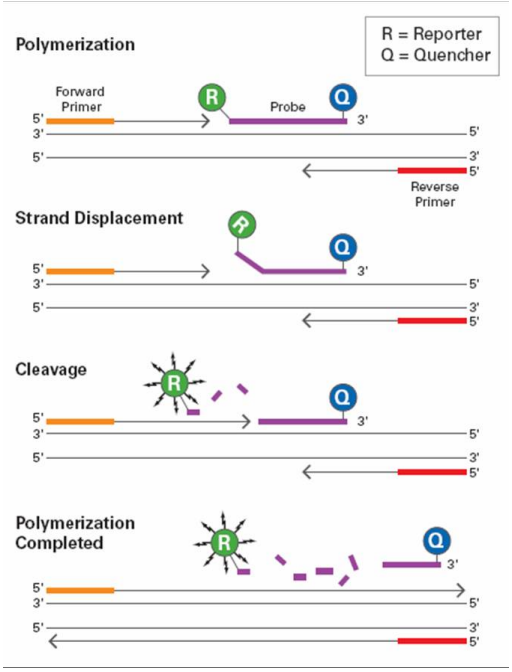

Figure 7. Principle of Taqman assay. The probe anneals downstream of the forward primer, and is cleaved by the DNA polymerase, increasing fluorescence emission from the reporter dye. This step will also remove the probe, allowing the completion of strand extension (Applied Biosystems).

1. 8 Restriction fragment length polymorphism (RFLP)

RFLP is a method used for the detection of variance across homologous DNA sequences, such as single nucleotide polymorphisms (SNPs). SNPs are variations in one nucleotide occurring throughout the genome, on average in every 1000th nucleotide (135) . Since only 3-5 percent of SNPs are present in the translated DNA sequences, most of the SNPs are silent mutations, but some are present in the coding DNA and may lead to protein modifications. SNPs may be associated with susceptibility to cancer, particularly in relation to genes involved in carcinogen metabolism, like CYP1A1 and GSTM1 (16, 46, 52, 135). RFLP is a simple approach to SNP analysis, involving restriction endonucleases to analyze PCR products where SNPs are localized (135). A PCR amplification of a DNA sequence is performed, including one or more SNPs of interest. The SNPs need to be associated with a restriction endonuclease site, and upon digestion of the PCR products, the fragments are separated by agarose or acrylamide gel electrophoresis. The fragment lengths will decide which SNP variant is included in the PCR product of interest. Uncut fragments represent the

homozygote of one variant, while the heterozygote shows two bands. The homozygote of the other genetic variant will be cut in both DNA strands, showing only one band, but shorter (runs further on the gel) than the uncut band. The PCR products may be sequenced to verify the results.

2. Aims of the study

Despite the improvements in cancer therapy, the five-year survival of patients with OSCCs remains low. Many patients, particularly in developing countries, are diagnosed with OSCCs at advanced stages due to lack of awareness of the symptoms related to this pathology and lack of access to health care facilities (52, 136). A better understanding of the molecular events possibly implicated in the development of OSCCs may enhance the development of better methods for early detection, predictive tools and therapeutic targets of this disease, which are essential to improve survival.

General aims:

Examine gene expression changes in OSCCs from high- and low-income countries to search for possible molecular biomarkers for OSCC and to identify genes associated with tobacco habits and other clinicopathological parameters.

Specific aims:

1. To study differential gene expression in OSCCs from Sri Lankan patients **(Paper I)**

2. To investigate differential gene expression in OSCCs from Sweden and the United Kingdom **(Paper II)**

3. To explore genetic alterations in OSCCs and OSFs from India and Sri Lanka **(Paper III)**

4. To study genetic variants of the *S100A14* gene in OSCCs and OSFs from Asian, African and European populations as a possible molecular biomarker for both OSCCs and OSFs **(Paper III)**

3. Materials and Methods

3.1 Patients (Papers I-III)

Primary samples of OSCC/OSF (**Papers I-III**) and their corresponding pair-wised normal controls (**Papers I and III**) were acquired from consecutive patients with previously untreated OSCCs/OSFs. After surgery, tissue samples (malignant and normal) were immediately submerged in the tissue storage and RNA stabilization solution, RNAlater™ (Ambion, Inc., Woodlands, TX, USA) and dispatched to the Department of Biomedicine at the University of Bergen, where they were stored at -20°C until RNA purification and microarray experiments were to be performed.

All tumors were staged following the 1987 UICC staging system, and had their histopathological diagnosis confirmed by some of the authors **(Papers I-III).** Histopathological diagnosis of the OSF cases **(Paper III)** was confirmed by two of the authors (RM/SOI) using formalin-fixed, paraffin embedded tissue sections stained with H&E. Diagnosis was done according to Pindborg *et al.* (18).

Ethical clearance:

All studies performed were approved by corresponding Committees for Medical Ethics at the University of Perideniya and at Moti Lal Nehru Medical College, Allahabad, India, King's College, London, UK and Uppsala University, Uppsala, Sweden.

3.2 Tissue samples and laboratory methods used

3.2.1 Patients and tissue specimens

Table 2. Overview of the tissue samples and laboratory methods used in Papers I-III

Tissue specimens	No	Laboratory procedures used	Paper(s)
OSCCs/Pair-wised normal controls (Sri Lanka)	15	cDNA microarray, RT-qPCR, RFLP	I, III
OSCCs (UK)	11	cDNA microarray, RT-qPCR	II
OSCCs (Sweden)	8	cDNA microarray, RT-qPCR, RFLP	II, III
OSCCs (Sri Lanka)	12	Array-CGH, RFLP	III
OSCCs (India)	12	Array, CGH, RFLP	III
OSFs (India)	6	Array-CGH, RFLP	III
OSFs (India)	54	RFLP	III
OSCCs (Finland)	45	RFLP	III
OSCCs (Norway)	24	RFLP	III
OSCCs (Sudan)	65	RFLP	III
NHOM, ODL & OSCCs (Sri Lanka)	21	IHC	III

NHOM, normal human oral mucosa; ODL, oral dysplastic lesions

3.2.2 RNA and DNA extractions

Total RNA was extracted from both tumor specimens (**Papers I and II**) and normal controls (**Paper I**) using TRIzol© 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®530 Life Science Spectrophotometer (Beckman Coulter, Inc., Fullerton, CA, USA) and by an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Genomic DNA (**Paper III**) was extracted using DNeasy Purification Kit (Qiagen Inc., Valencia, CA, USA). DNA quantity and quality were determined using Nanodrop (Thermo Fischer Scientific).

3.2.3 cDNA synthesis, DNA labeling, hybridization and scanning (Papers I-III)

Synthesis and labeling of the cDNA was carried out using Fairplay Microarray Labeling Kit (Stratagene, La Jolla, CA, USA), following the manufacturer's instructions. Synthesized cDNA was labeled with Cy™3 (control cDNA) and Cy™5 (tumor cDNA) monoreactive dyes (Amersham Biosciences, GE Health Care), and samples were hybridized to the human oligonucleotide microarrays containing 34.580 oligonucleotid probes **(Paper I)** or 21.521 oligonucleotides **(Paper II)** printed on Corning Ultra GAPS slides at the Norwegian

Microarray Consortium (www.mikromatrise.no). Labeled cDNA **(Paper I and II)** was hybridized on the Ventana Discovery® XT System (Ventana Medical Systems Inc., Tucson, AZ, USA) according to the manufacturer's protocols.

Genomic DNA **(Paper III)** was labeled with Cy™3 (control DNA) and Cy™5 (tumor DNA) and hybridized to microarrays containing 4.549 BAC and PAC clones representing the human genome at ~1 Mb resolution, as well as the minimal tiling- path between 1q12-q25. Labeled DNA hybridization was performed using an automated hybridization station GeneTAC/HybArray (Genomic Solutions, Ann Arbor, MI, USA). cDNA microarray and DNA microarray slides were scanned by Agilent DNA Microarray Scanner BA (Agilent Technologies, Palo Alto, CA, USA), and the microarray data was 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.

3.2.4 Quantitative Real-Time RT-PCR (Papers I and II)

To validate gene expression profiles for selected candidate genes, real-time quantitative RT-PCR was performed for selected genes (9 in **Paper I,** 3 in **Paper II**). Aliquots of the same RNA (200-300 ng) used for the microarray hybridization was also used for synthesis of the cDNA, performed with High Capacity cDNA Archive kit (Applied Biosystems, Foster CA), following the manufacturer's instructions. Real-time PCR was performed with probes for each gene using the ABI 7900 HT (Applied Biosystems) and 384 well optical plates (ABI). Each reaction contained 1µl cDNA, 5µl 2xTaqMan Universal Master mix (Applied Biosystems), 0.5µl Taqman AOD probe and H_2O to a final volume of 10 µl, and was run in triplicate. Cycling parameters were 95ºC for 10 min, followed by 40 cycles of 95ºC for 15 sec and 60ºC for 1 min. Serial diluted standards were run on the same plate and the relative standard curve method was used to calculate gene expression. β-actin was used as an endogenous normalization control to adjust for unequal amounts of RNA.

3.2.5 Tissue specimens and immunohistochemistry (Paper III)

Immunohistochemistry (IHC) is a technique applied to identify proteins in tissues, based on antigen localization detection by antibody-antigen binding, visualized by fluorescence, enzyme-labels, colloidal gold particles or other methods (137, 138). Usually, the antibodies in use are derived from another species than the tissue to be examined. Rabbits and mice are

widely used for antibody production, but also other mammals are used, like goat, guinea pig, cow, swine, rat and camel (138). The IHC method was first introduced in the 1940s, and has evolved to be an important and widely applied technique in many medical research laboratories as well as in clinical diagnostics (137).

IHC analysis of the S100A14 protein was performed on 21 archival formalin fixed, paraffin embedded tissue sections of OSCCs (representing 10 of the 15 patients from Sri Lanka) including oral dysplastic lesions and normal human oral mucosa using Autostainer universal staining system (DAKO-USA, Carpinteria, CA). Antigen retrieval was done by microwave treatment in Tris-EDTA buffer, pH 9.0 (DAKO). After blocking with 3% BSA in TBST, sections were incubated with rabbit polyclonal anti-human S100A14 primary antibody (10489-1-AP, Proteintech, Chicago, IL, USA, 1:500 dilutions) for 1 hr at room temperature. After washing, anti-rabbit secondary antibody conjugated with horseradish peroxidase labeled polymer (EnVision System, DAKO) was applied. Presence of antigen was visualized by staining with 3, 30-diaminobenzidine (DAKO), counterstained with Harris hematoxylin (DAKO) and mounted with EuKit mounting medium. Sections incubated with 3% BSA instead the of primary antibody served as negative controls.

3.2.6 Evaluation of the IHC

Tissue sections were examined using a light microscope for S100A14 expression. IHC evaluation was mainly focused on the invading islands of the OSCC specimens. S100A14 staining was semi-quantitatively evaluated by manually counting the cells (at least 500 cells were counted in 3 representative areas, at 400X magnification) expressing either membranous or mixed membranous/cytoplasmic S100A14. Based on the number of positive cells with respect to sub-cellular localization of the S100A14, OSCCs were categorized into 3 groups with *low* (0-9% positive cells), *moderate* (10-49% positive cells) and *high* (50-100% positive cells) scores (92).

3.2.7 Restriction fragment length polymorphism (RFLP)

To search for SNPs in the *S100A14* gene, we performed PCR with two different sets of primers within the *S100A14* gene searching for two SNPs, 461G>A (rs11548103) and 1545A>T (rs11548102). All available DNA samples from the cases used for array-CGH in **Paper II** were included in the RFLP study. For comparison, and further validation, we included archival OSCC DNA samples from Sudan (*n=65*), India (*n=60*), Finland (*n=47*) and

Norway (*n=24*). In addition, a collection of archival material from patients diagnosed with oral submucous fibrosis from India (*n=43*) were included. The PCR products were digested with restriction enzymes *Kpn* I (461G>A), *Mnl* I (1545A>T) and separated by 3% agarose gel electrophoresis. The agarose gel was stained with 1X GelRed Nucleic Acid Staining solution (Biotium, Hayward, CA, USA).

3.2.8 DNA sequencing

10 PCR-products of the 461A>G SNP were selected for verification by sequencing. The PCR reaction was performed as described in the previous section. The products were run on a 3% agarose gel, excised from the gel and purified with Qiagen gel extraction kit (Qiagen Inc., Valencia, CA, USA) and ligated into the pCR®2.1-TOPO® vector (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. After ligation, the vector with insert was transformed into XL10-Gold Ultracompetent *E.coli* cells (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocol, plated on LB^{amp} plates and grown overnight. Clones were picked and grown overnight in 10mL LB^{amp}, and plasmids were extracted using the Promega plasmid prep kit (Promega, Madison, WI, USA. The sequencing reaction was performed at the DNA sequencing facility at the Department of Molecular Biology, University of Bergen (www.seqlab.uib.no).

3.3 Statistical analysis

For **Papers I and II**, J-Express software package (version 2.6; www.molmine.no) was used to analyze the gpr-files produced by GenePix Pro, and the pre-processing was performed by filtering and normalization. In **Paper I**, each array was first pre-processed separately by performing the following steps: Spots flagged by Genepix ("bad", "absent" or "not found", -100, -75 and -50, respectively) were filtered; 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. Global lowess normalization was applied to all values left after the filtering step. Thereafter, all in-array replicate spots were merged by a median statistics and inserted into a gene expression matrix. Genes for which more than 50% missing values (missing in more than 50% of the patients) were removed. Missing values were set to zero, thus eliminating their contribution for recording up- or down-regulation status of a gene in

tumor compared to 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 tumor and normal control, the relative difference in gene expression d(i)=M(i)/(SE(i)+s) was used where M(i) is the mean log ratio for gene i, SE(i) is the standard deviation of the gene's log ratios and s is an added constant for all genes. This means that d(i) is a student's t-statistics with a fudge factor s which corrects for underestimated variances resulting in a higher weight to high average fold change compared to low variance that can be justified by noisy nature of microarray experiments. In our case, and as suggested by Efron *et al.* (139), s is set to a 90 percentile of all gene standard errors SE(i).

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

For the tumor expression matrix, we selected genes with an s-score above 1.0, and obtained 461 genes as differentially expressed either between the tumors and normal controls or alternatively, as a result of the dye effect. To remove genes affected by a dye effect, we removed all genes for which array vendor had reported a possible dye effect, and also genes with s-score above 0.5 in the dye swap expression matrix, resulting in 263 genes that we believe are differentially expressed between tumors and normal controls. We further performed a permutation experiment and generated 1000 permuted matrices. Each permuted matrix contains the 12034 genes in the original unfiltered data set and the permutation was performed by flipping the sign of each column with a probability 0.5. Averaged over the 1000 permuted matrices, our analysis generated 0.47 genes per matrix with a d-score above 1.0 resulting in a false discovery rate of 0.001.

A pre-processing of each array was also performed in **Paper II**, by removing controls and spikes, as well as flagged and empty spots. Multiple spots for the same gene were

combined and represented by the median intensity value. The Global lowess normalization method was applied to each array individually. Genes with more than 40% missing values (missing in more than 40% of the patients) were removed. Remaining missing values were imputed with KNN input (K=10), the data were scale normalized (inter-array normalization as described by Yang *et al.* (140)) and finally genes showing little variation across the samples (s.d. less than 0.5) were removed. This resulted in log-ratios for 2439 genes across the 19 samples studied. Significance analysis of microarrays (SAM) was performed with the 2439 genes, using two classes (UK, Sweden; un-paired, with 1000 permutations) to find genes differentially expressed between the Swedish and UK samples. The percentage of genes falsely identified as differentially expressed, False Discovery Rate (FDR), was set to less than 4 percent, resulting in 42 genes.

To search for over-represented Gene Ontology terms (**Papers I and II**), Gene Ontology (GO) Directed Acyclic Graph (DAG) analysis were used. Genes related to the same biological pathway were found by performing a KEGG (Kyoto Encyclopedia of Genes and Genomes, www.KEGG.com) analysis, using Fisher-Irwin exact test to find statistically significant pathways (**Paper I**). Fischer-Irwin exact test was also used to analyze data from RFLP (**Paper II**). Results from real-time quantitative PCR were analyzed by the Mann-Whitney U-test (GradPad Software Inc, La Jolla, CA, USA) (**Papers I and II**)

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

For array-CGH data (**Paper III**), gpr-files produced by GenePix were further processed by the MATLAB tool M-CGH, performing filtering and log2 normalization of CGH ratios, and linking the clones to their corresponding genomic information in the Ensembl database (www.ensembl.org). Significant copy number changes common for the two populations were only included when found in a minimum of 20% of the Sri Lankan and Indian samples. Genetic alterations were investigated separately for the two populations, to find DNA copy number changes exclusive for Sri Lankan or Indian samples. Alterations in one population were only considered as exclusive when no changes were seen in the other population.

4. Results and general discussion

During recent years, microarray-based technologies have become commonly used techniques for identifying gene expression and chromosomal alterations in human cancers and other pathological conditions. These high-throughput technologies enable genome-wide analysis of changes in gene expression or chromosomal deletions/amplifications in the pathological samples to be studied. In the search for possible molecular biomarkers for OSCCs, a series of studies were carried out in the present work where both cDNA microarrays and array-CGH were applied to examine possible changes in gene expression and DNA copy number alterations in OSCC samples compared to their pair-wised normal controls and between OSCC samples from different populations. The results showed significant findings that will be presented and discussed as follows:

4. 1 Gene expression profiles and chromosomal deletions/amplifications (Papers I-III)

In **Paper I**, cDNA from OSCCs and their pair-wised normal controls obtained from patients from Sri Lanka were hybridized to 35k oligonucleotide microarrays, and 262 genes were found as differentially expressed in the OSCCs compared to normal controls (**Tables IIIA and IIIB in Paper I**). Among the 262 genes, 189 were found as up-regulated (including *KRT17*, *COL4A1*, *MMP1*, *MMP3*, *HRAS* and *JUN*) and 73 genes were found as down-regulated (including *KRT19*, *NDRG2*, *S100A1*, *EEF1G* and *WDFC2*). Among the 262 genes, 191 were of known functions (related to cell signaling, cell growth, cell adhesion, angiogenesis and other processes important for oral tumorigenesis) and 71 were of unknown function (several were without any assignation to a gene symbol). Interestingly, 66 (35%) of the known genes have previously been reported in OSCCs, according to the Cancer Genome Anatomy Project (CGAP) and gene expression studies performed on OSCCs**.** Further, hierarchical clustering was performed with the expression profile of the 262 genes found as differentially expressed (**Figure 1 in Paper I).** The samples were separated in a number of sub-groups, with the exception of samples 1 and 11. Sample 1 was diagnosed as a verrucous carcinoma - a tumor related to ST use that is usually less aggressive compared to other OSCCs. Sample 11 was an advanced OSCC. A large subgroup of the samples with tumors predominantly of stage 3 and 4 was observed, which might suggest that advanced tumors have

a common gene expression profile compared to tumors of stage 1 and 2. Interestingly however, one of the stage 2 tumors clustered in a subgroup with tumors of stage 3 and 4, suggesting a gene expression profile in this sample that is similar to advanced tumors (**Figure 1 in Paper I**). **In Paper II**, cDNA from OSCCs samples from Sweden and UK were hybridized with human universal reference RNA to 21 k oligonucleotide microarrays, and 2439 genes were found to be differentially expressed. Significance of Analysis of Microarrays (SAM) was performed to find differentially expressed genes between Sweden and UK OSCCs, resulting in 42 genes. Based on the 2439 genes found, we performed hierarchical clustering of the samples studied from the two countries and the results showed a tendency for most of the patients grouping based on country of origin (**Figure 1 in Paper II**). All subgroups included tumors of different stages. One case from UK and one case from Sweden did not group with any of the other cases. Interestingly, the case from UK is the only case among the UK patients with a history of ST use (in addition to smoking), and the case from Sweden is the only sample with a distant metastasis (**Figure 1 in Paper II**). In **Paper III**, comparative genomic hybridization analysis using array-CGH was performed in OSCCs cases from Sri Lanka and OSCCs/OSFs samples from India. This resulted in 95 chromosomal regions found as deleted or amplified with 14 regions being deleted in samples from both populations and 14 regions with chromosomal gains. Chromosomal regions deleted in one population and amplified in the other were also found, resulting in 66 regions predominantly found as deleted in Sri Lankan and amplified in the Indian samples. The chromosomal regions found as deleted in both populations included among others 1q21, 1q23, 8q22 8q24, 12p13, 13q12 and 20p11. Amplificated regions included chromosomes 1p36, 6p21, 7q36, 11q13, 15q23, 18p11 and 19p13. Of the 66 regions with either DNA copy number increase or decrease in one of the two populations were 1p36, 7p22, 7q22, 11q13, 12q24, 16p11, 17p13, 17q21, 17q25, 19p13 and 22q13. Several of these chromosomal regions are common DNA copy number alterations in OSCC (16, 54). The Ensembl (http://www.ensembl.org/index.html) and GeneCards® databases (http://www.genecards.org/) were applied to identify genes of biological significance located in chromosomal regions with copy number changes (34 genes in the deleted chromosome regions, 30 genes in amplified regions and 285 genes in amplified or deleted chromosomal regions in the two populations). Among the genes found in regions with a decrease in DNA copy number were *IVL*, several S100 gene family members, including *S100A14*, *ANGPT1*, *WISP1*, *NDRG1*, *CASP4* and *CASP5*. Regions with increased DNA copy number represented genes such as *NOTCH 4*, *XRCC2*, *MMP15*, *PRKCA*, *BCL2* and *CD70*. Genes localized in chromosomal parts deleted or

amplified in either of the two population included *MTOR*, *ECM1*, *BAK1*, *AZGP1*, *CAV1*, *CAV2*, *PLAU*, *TNC*, *VEGFB*, *BAD*, *CCDN1*, *FGF3*, *FGF4*, *MMP14, EIF4A1*, *TP53*, *TIMP2* and *TGFB1*. These genes are listed in **Tables 2A** and **2B and Supplementary table S1** in **Paper III**. When combining the results from all three studies for the samples studied from Sri Lanka, India, Sweden and UK, a total of 18 genes were found to be represented (**Papers I-III**). These genes included caveolins, plasminogen activator, fatty acid synthase, hyaluronan synthase 3, tenascin C and protein arginine methyl transferase. When including previous work from the group where samples of OSCCs from Norway and Sudan were studied using microarray gene expression analysis, a total of 72 genes (**Table 3 in Paper III**) were found to be represented in more than one study, including *CCND1*, fibroblast growth factors, *TP53*, S100 gene family members, matrix metalloproteinases and *TGFB1*.

4.1.1 Gene expression profile for selected genes (Papers I-III)

4.1.1.a Caveolins, matrix metalloproteinases, extracellular matrix, proteases and oncogenes

Many of the common genes found are related to tumor development, progression, invasion and metastasis, and some are possible biomarkers for OSCCs. Caveolin 1 (*CAV1*) and 2 (*CAV2*), genes (detected in **Paper I and III)**, are structural proteins of the caveolea, 50-100 nm flask shaped invaginations of the plasma membrane, distributed in a wide variety of cells, abundant in adipocytes, endothelial cells, pneumocytes and smooth muscle cells (141, 142). A third caveolin (*CAV3*) has also been identified, not as ubiquitously expressed as caveolin 1 and 2 (141). Caveolin 1 and *2* are co-expressed and form a stable hetero-oligomeric complex. Caveolin 1 and 3 are able to form caveolins alone, while caveolin 2 is not. The functional roles of the caveolae and caveolins are connected to cellular processes like endocytosis and signal transduction. Caveolins (particularly *CAV1*) and caveolae are associated with cancer development and progression (141, 142). *CAV1* has been suggested as a tumor suppressor due to its down-regulation in malignancies like breast, lung and colon cancer. However, in some cancers like esophageal SCCs, OSCCs, prostate and breast an increase in expression has been observed (141, 143). These findings suggest several functional roles for caveolin 1 depending on the cell type and tissue investigated. *CAV2* has not been studied as extensively as *CAV1*, but the expression pattern of this gene in cancer seems to follow a similar pattern as *CAV1*. Our findings suggest an increase in expression of *CAV1* and *CAV2* in OSCCs, consistent with previous studies performed with OSCCs and OSCC cell lines (144-146). Also, in a previous

proteomics study including OSCCs from Sudan and Sri Lanka, we found increased expression of caveolin 1 (147).

MMPs (Matrix metalloproteinases) are a family of more than 20 proteolytic enzymes responsible for remodeling of extracellular matrix components. MMPs are normally not expressed in healthy tissue. Their precursors are produced and can be found in many cells, but the expression of MMPs is strictly regulated. However, MMPs are upregulated in tumor tissue, being involved in processes such as cell proliferation, cell migration, invasion, metastasis and angiogenesis (148, 149). **Papers I and III** included four MMPs; *MMP1*, *MMP3*, *MMP14* and *MMP15*. *MMP1* and *MMP3* were also reported in another OSCC study performed by our group (81). *MMP1* (Interstitial collagenase) encodes a protein that breaks down collagens I, II, III, IV and V. MMP3 (Stromelysin-1) also breaks down the same collagens, in addition to IX, X, XI and elastin. *MMP14*, a cell surface MMP (MT1-MMP), is involved in activation of pro-MMP2 by binding to *TIMP2* (Tissue inhibitor of metalloproteinase 2). *MMP15* (MT2-MMP) is also a cell surface MMP. Substrates for this enzyme are collagenes I to III, elastin, fibronectin, gelatins and laminin (149). *MMP1*, *MMP3* and *MMP14* has previously been reported to have elevated expression in oral cancers (97, 114, 148-152). Collagen IV, a basement membrane constituent is one of the substrates for both MMP1 and MMP3 proteins. Up-regulation of *MMP2* and *MMP9* seems to occur frequently in oral cancer (148, 149, 153, 154). Our earlier work showed differential expression of *MMP2* and *MMP9* between OSCCs from Sudan and Norway (99), and another study demonstrated involvement of *MMP1* and *MMP9* in invasion and metastasis in OSCC cell lines (92). *COL4A1* (Collagen IV, alpha I) is a subunit of collagen IV, found as up-regulated in **Paper I** and our previous work (81). Collagen IV has a possible implication in OSCC as it has been found as surrounding cancerous cells in ECM (extra cellular matrix), and that it is interrupted and discontinuous in invasive carcinoma (148). Collagens are also associated with oral submucous fibrosis, a premalignant condition mainly caused by BQ chewing. Collagen IV has not been directly connected with OSF, but Tsai *et al.* (152) reported *COL4A1* as up-regulated when investigating mRNA profiles of Taiwanese BQ chewers diagnosed with OSCC. *COL1A1*, another gene reported as up-regulated in **Paper I**, is connected with a risk of developing OSF for individuals carrying a specific *COL1A1* genotype (155). This might indicate that expression of COL4A1 and COL1A1 in OSCCs are connected with BQ chewing. In contrast, Cheong *et al.* (114) studied OSCCs from smokers and BQ chewers, demonstrating elevated expression for COL4A1 and COL1A1 in both groups, indicating that both etiological factors can affect these genes. Additionally, these

genes were found to have altered expression in previous OSCC studies (97). In **Paper III**, several collagens were detected in amplified or deleted chromosomal regions; *COL5A3*, *COL6A1*, COL6A2, *COL11A2*, *COL13A1*, *COL15A1* and *COL16A1*. *SPARC* (Secreted protein and rich in cysteine) is also an ECM protein located in the basement membrane, found as up-regulated in **Paper I** and in our previous work (80, 81). This gene has been demonstrated to have elevated expression in some malignancies and of decreased expression in others (156). In oral cancers however, *SPARC* is up-regulated and is associated with tumor progression, metastasis and survival (97, 157, 158). Another ECM protein, tenascin C (*TNC*), is expressed in tumor tissue and promotes tumor cell proliferation, invasion and angiogenesis (159, 160). Tenascin C interacts with ECM molecules such as fibronectin and perlecan, as well as a number of cell surface receptors, including integrins. TNC has been associated with tumor progression and poor prognosis (159, 160), but this is not the case for OSCCs (148). In **Paper I and III**, *TNC* was detected as up-regulated or located in an amplificated chromosomal region, and has been reported to have elevated expression in several studies performed with head and neck SCCs (97, 114). TNC protein might be implicated in up-regulation of *MMP9* and hereby invasion and metastasis (92, 148, 161).

Proteases are important for cancer cells to be able to invade the surrounding tissue and to enter blood and lymph vessels, mostly performed by *MMPs*. There are also other enzymes responsible for ECM protelysis, like the urokinase plasminogen activator (*PLAU*). This protease has a restricted substrate specificity, activating plasmin from its inactive form plasminogen (162). When activated, plasmin is able to degrade ECM proteins and also to activate MMP precursors, like pro-MMP*3*. Additionally, plasmin has the ability to release or activate growth factors, such as *TGF*β (162). PLAU (also named uPA) binds to its receptor PLAUR (also known as uPAR or CD87), forming a complex enhancing proteolysis, cell proliferation, migration and modulating cell adhesion. PLAU and PLAUR are believed to play a critical role in tumor invasion and metastasis (162, 163). *PLAU* was detected in **Paper I** and **III** as well as in our previous studies (97, 114, 152). The involvement of extracellular matrix components in oral tumorigenesis is important for tumors to invade the connective tissue, to migrate and metastasize to distant sites. The genes described here may act as possible therapeutic targets and prognostic tools for OSCCs. Clinical trials on inhibitors of MMPs have been performed, although promising, there are yet no established successful therapies. SPARC and PLAU have been described as having a prognostic value for OSCC (158, 164).

In Paper II, we detected 42 genes as differentially expressed in OSCCs from Sweden compared to UK. Four of these genes (*APOL3*, *PRMT1*, *FASN* and *HBA1*) were also found in **Papers I and III**. *HBA1* (Hemoglobin, alpha 1) has no clear role in tumorigenesis. Apolipoprotein 3 (*APOL3*) belongs to a cluster of six apolipoprotein L genes, APOL1-6, proteins involved in lipid transport and metabolism. *APOL3* has been shown also to have a role in inflammation and apoptosis (165, 166). Johanneson *et al.* (167) demonstrated that genetic variance of *APOL3* is connected to prostate cancer susceptibility. *PRMT1*, protein arginin methyltransferase 1, was reported in **Papers I-III**. *PRMT1* encodes a protein which methylates arginin residues in proteins. There are several substrates for PRMT1, one of the major substrates are the histones, in particular H4, which is dimethylated at arginine 3 by PRMT1 (168). Methylation of the histones is one of the epigenetic events which regulates gene expression (169), and hereby plays an important role in cancer development (170). Arginine methylation modulates pathways such as signal transduction, epigenetic regulation and DNA repair pathways, and the involvement of PRMT1 in cancer progression is highly probable (164). A function for PRMT1 as a co-activator of p53-mediated transcription has also been demonstrated (171). Fatty acid synthase (*FASN*) catalyzes the synthesis of long chain saturated fatty acids. In cancerous conditions, *de novo* synthesis of fatty acids is elevated, due to an increase in the need of lipids in the growing cancer cells (172). High levels of FASN expression have been found in many human tumors, including oral cancers (172). Tumor- related FASN expression is regulated by growth factors such as EGF and its receptor *EGFR*, *ERBB2*, and steroid hormones (androgen, estrogen and progesterone) and their respective receptors. These signals activate the PI3K, Akt and MAPK signaling pathways, stimulating *FASN* translation through SREBF1. Low pH and hypoxia will also induce FASN expression by activating SREBF1. FASN mutually stimulates ERBB2 signaling, increasing tumor cell proliferation. There are also transcription factors involved in *FASN* regulation, as well as posttranslational regulation (172, 173). The role of FASN in tumor growth and survival provides a possible therapeutic approach. FASN inhibitors have been shown to suppress tumor growth, but further studies on the mechanisms of FASN in tumor growth are needed before a therapeutic strategy can be applied (173).

The oncogene *FOS* (FBJ murine osteosarcoma viral oncogene homolog) was detected in **Paper II**. The Fos protein can dimerize with c-jun, a member of the Jun family, to form the transcription factor *AP1* (Activator protein 1) (174). The AP1 complex regulates genes that are responsible for cell proliferation, invasion and metastasis. Other members of the Fos and Jun family also form AP1 proteins, and Jun family members can form jun-jun homodimers

(174), although these homodimers are less stable and weaker DNA-binding activity than the heterodimers. c-Fos is frequently over-expressed in a number of malignancies, and has oncogenic functions. However, recent studies have revealed tumor suppressor activity of c-Fos, suggesting a dual role for this protein in cancer (175, 176). Previous studies of gene expression in head and neck SCCs have found this gene to be both up- and down-regulated (97)

4.1.1.b S100 A gene family members

Altered gene expression/DNA copy number changes were reported in S100 genes in **Paper I** (*S100A1*) and **III** (*S100A1-9*, *12-14* and 16). In **Paper III**, the chromosomal 1q21 region harboring *S100A14* was found to be deleted, while in a previous CGH study performed on OSCCs from Sudan and Norway, an amplification of the 1q21 region has been reported (81). An up-regulation of *S100A2* was also reported in our previous work (80), as with down-regulation of *S100A4, S100A6, S100A8 and S100A14* studying expression of S100 genes in OSCCs (78). An in-depth study on gene expression profiles of *S100A14* and 18 other S100 gene family members in OSCCs from Sudanese patients showed down-regulation of *S100A14*. S100 proteins are calcium-binding proteins with a broad diversity of both intra- and extracellular functions, such as motility, cell growth, differentiation, transcription and secretion (68). S100 proteins are associated with human pathological conditions, including psoriasis, rheumatoid arthritis, neurodegenerative disorders, cardiomyopathies and a number of malignancies (63, 64, 69, 71). *S100A1* plays a role in different pathological conditions, and there are numerous target molecules for S100A1 (177). Some of the biological functions connected to the S100A1 proteins are Ca^{2+}-regulation, phosphorylation, cell growth, apoptosis and cytoskeleton organization (177). Its role in cancer is not clear, but an interaction with the metastasis-inducing S100A4 protein was shown in rat tumor derived cell lines, where S100A1 had an inhibitory effect on S100A4-inducing metastasis (178). S100A2 is a possible tumor suppressor, down-regulated in several malignancies (64). In oral cancers, S100A2 has been shown to interact with p53 (75), and down-regulation may suggest this protein as a tumor suppressor also in OSCCs (82, 83, 179). Increased expression of S100A4 has been associated with a number of cancers (64, 71), and has been suggested to play a role in metastasis and invasion. An interaction between p53 and S100A4 has also been demonstrated (64, 71). In OSCCs, S100A4 seems to be involved in metastasis and tumor progression (180, 181). S100A7 expression was first identified in psoriatic keratinocytes, alternatively named psoriasin. Expression of S100A7 in different malignancies has been demonstrated, and also in

OSCCs (71, 85, 88). It seems that S100A7 expression is higher in early stage carcinomas than in moderate or poor differentiated tumors (85, 86, 88). An association between S100A8 and HPV18-infected OSCC was reported by Lo *et al.* (89), suggesting an important role for this protein in OSCC tumorigenesis after HPV18 infection. Differentially expressed S100A8 in OSCCs compared to normal mucosa was detected in another study, also showing differential expression of S100A9, 11 and 13 (90).

The biological functions of the S100A14 protein are not yet understood, but it is clear that it is involved in different cancers, as S100A14 has altered expression in several malignancies. It was over-expressed in breast, ovarian and uterine tumors, and down-regulated in kidney, colon, rectal, esophageal and, as mentioned above, in OSCCs (78, 79, 91). Chen *et al*. (77) studied different genetic variants of *S100A14* in esophageal SCCs, and demonstrated that transcription of *S100A14* is regulated by the tumor suppressor protein p53, and that one of the genetic variants (461G>A) may correspond to a p53-binding site (77), and suggested that S100A14 is a putative tumor suppressor for esophageal SCCs. Further, recent work by our group showed that over-expression of S100A14 in OSCC-derived cells harboring wild-type p53 induced G1-arrest and hereby suppressing cell proliferation (76). Additionally, S100A14 was found to regulate the invasive potential of two OSCC cell lines, involving down- or up-regulation of matrix metalloproteinases *MMP1* and *MMP9*. When over-expressing S100A14, MMP1 and MMP9 were found to have decreased expression, thus inhibiting invasion of OSCC cells (92). In this experiment *TNC* was down-regulated upon over-expression of S100A14, supporting our findings of these genes playing an important role in OSCC development. S100A14 may be favoring accumulation of p53. *TP53*, the "guardian of the genome", is involved in many cellular processes, preventing abnormal cells to divide and by inducing apoptosis, DNA repair, cell cycle arrest and inhibition of angiogenesis. *TP53* mutations occur in a great number of human cancers, including more than 50% of OSCCs, leading to a dysfunctional protein. More than 70% of the mutations affect the DNA-binding site. p53 inactivation also occurs through alterations in genes encoding proteins involved in regulation of this protein, like *MDM2*, or viral protein binding (50, 56). Tobacco products contain several carcinogens responsible for causing mutations in *TP53*. For ST, the main tobacco habit among patients in this study, TNSAs like NNK and NNN are the strongest carcinogens present (27, 31). Previously, we showed a relationship between use of oral snuff (*toombak*) in Sudan and *p*53-mutations (57). Also, BQ chewing may be responsible for *TP*53 mutations as demonstrated by Chiba *et al*. (182) and Ralhan *et al*. (183). Alterations in *TP*53-hosting chromosome part 17p13.1 were found in **Paper III**, deleted in the Sri Lankan tumors

and amplified in the Indian tumors. Our previous study involving OSCCs from Sudan and Norway reported this region as amplified (81). Findings in the present work suggest that there are several common pathways for OSCC development regardless of differences related to life-style and ethnicity. Among the genes implicated in OSCCs from populations world-wide, there are possible biomarkers for OSCC development and progression in general, and also putative biomarkers related to tobacco habits. Our findings have presented genes that are differentially expressed in different oral cancer samples, and genes that may be up-regulated in one type of cancer and down-regulated in another. A gene that is up-regulated in a cancer cell line may not exhibit increased expression *in vivo*, since the tumor cells also respond to signals from their microenvironment. The fact that the expression of a gene is changed in one type of tissue compared to another does not necessarily mean that the expression of its gene product has changed. Genes may undergo post-transcriptional or translational regulation, affecting mRNA transcripts or proteins. For microarray experiments however, the results may depend on the source of samples and sample processing, as there is always a risk of bias due to the experimental procedures. Reproducing results can be difficult in two different laboratories, even with aliquots of the same samples and following the same procedure, since human handling could affect the results. For this reason, microarray experiments must always be carefully planned and interpreted. When comparing results from different microarray experiments, it is important to keep these considerations in mind.

4. 2 Common biological pathways

Different pathway databases were searched (KEGG pathways (www.kegg.com), Nature Pathway Interaction Database (NPID) (http://pid.nci.nih.gov/index.shtml) and Millipore Pathways (http://www.millipore.com/pathways/pw/pathways)) to find common pathways for the genes discussed in section 4.1.1. The Millipore pathway Cell adhesion ECM remodeling includes a number of genes, such as *MMP*s, *PLAU*, *EGFR*, *SPARC* and collagens (Figure 8). Many of the same genes are represented in Focal adhesion, a KEGG pathway including ECM-receptor interaction and Cytokine-cytokine receptor interaction. *PLAU* and *MMP*s are represented in NPID by Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling. To date there are only a few connections of *S100* genes with known biological pathways. *S100A2* has been related to direct p53 effectors (NPID), a p53 signaling pathway. This pathway also includes *PRMT1*, *MMP2* and *CAV1*, establishing a link between the S100 genes and the genes described in section 4.1.1.a.

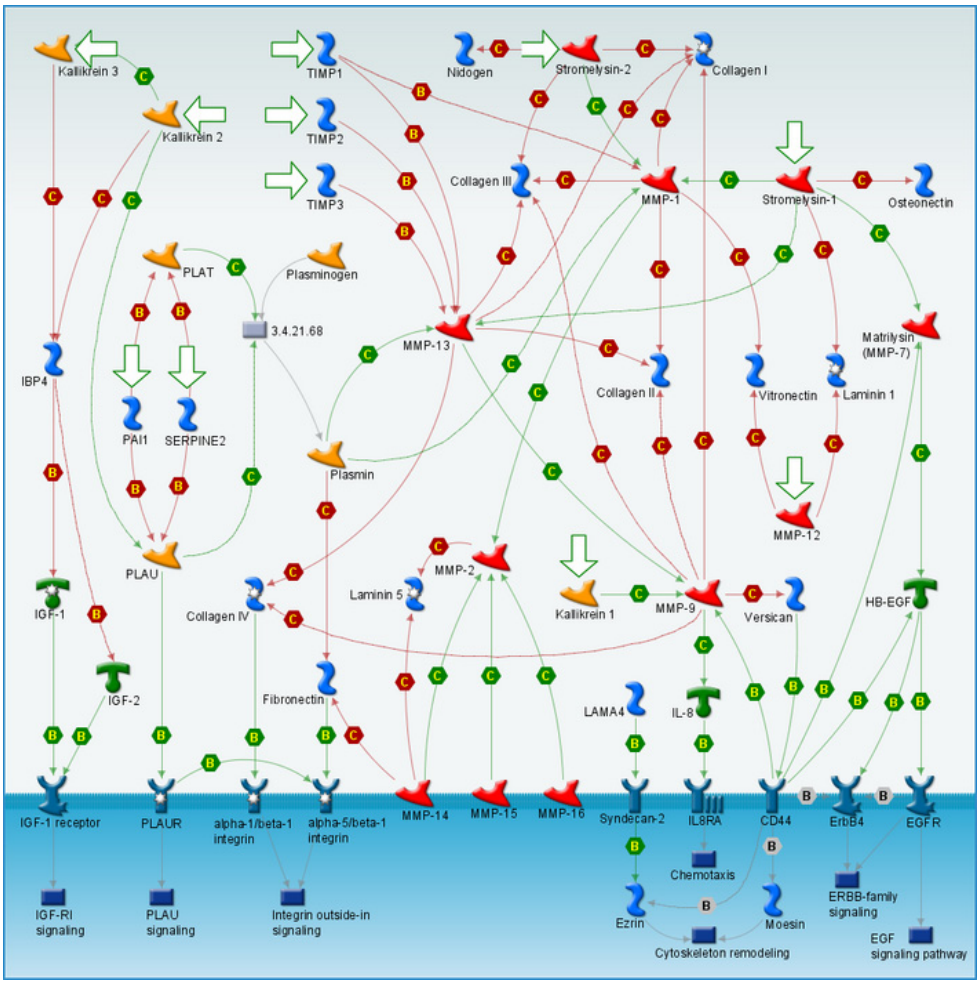

Figure 8. Cell adhesion ECM remodeling. This pathway involves several genes found in this work, such as MMPs, collagens, PLAU, SPARC and EGFR. Remodeling of the ECM is important for many tumorigenic processes such as angiogenesis, cell proliferation, motility and adhesion. (www.millipore.com).

4. 3 Chromosomal alterations in OSCC compared to OFS samples (Paper III)

By searching for chromosomal regions with differential changes in OSCC compared to OSF, very few alterations affecting OSCC and not OSF were observed (**Table 5 in Paper III**). There were no observations of DNA copy number changes affecting OSF exclusively. Hierarchical clustering of the Indian patients showed that the OSF samples grouped separately (**Figure 1 in Paper III**). Some of the genes located in the altered regions in both OSCC and OSF included *EGFR*, *SULF1*, *EXT1*, *MTSS1* and *MLNA*. *EGFR* (epidermal growth factor receptor) is a well known tumor-promoting gene, being a receptor for EGF and other growth

factors. Over-expression of the EGFR protein is a common finding in cancers, associated with poor survival in some malignancies, including OSCCs (59, 61). Findings of *EGFR* up-regulation have also been related to premalignant lesions. Chen *et al*. (184) demonstrated a connection between EGFR over-expression and poor survival in OSCC patients with a history of BQ chewing. The *EGFR* gene has been suggested as an important therapeutic target for OSCCs and other malignancies. Different therapeutic strategies for *EGFR* in OSCCs have been approached, with promising results (59, 185). Our results, however, contradict earlier findings, as *EGFR* was localized in chromosome 7p11.2, a region found as deleted in OSCC, and not found in OSF.

Sulfatase 1 (*SULF1*) is a member of a family of esterases responsible for hydrolyzing ester bonds in different substrates. *SULF1* encodes an extracellular heparan sulfate endosulfatase, suspected to have a tumor suppressor function, being down-regulated in many cancer cell lines (186). Sulfatase 1 has been implicated in inducing apoptosis, inhibition of heparin binding growth factors, angiogenesis inhibition and other tumor suppressing functions (186). However, further studies are needed to elaborate the tumor suppressor role of SULF1.

EXT1, exostosin 1, encodes a glucosyltransferase required for heparan sulfate biosynthesis. This protein is involved in the autosomal bone disorder called hereditary multiple exostoses (HME), manifesting with benign bone tumors. Malignant transformation of these benign tumors are seen in 1-2% of HME patients (187). Reduced expression of EXT1 due to mutations is seen in HME, and loss of heterozygozity in transformed malignant tumors, suggesting tumor suppressor activity for EXT1. Reduced expression of EXT1 in other cancer cells have also been reported (187). Both *SULF1* and *EXT1* are localized in chromosomal regions found to be deleted in this study. The fact that the majority of DNA copy number alterations were common both for OSCCs and OSFs supports the idea that the majority of the chromosomal changes occur early in OSCC development. However, the OSF sample size is too small to enable us to draw any conclusion from these results.

4. 4 S100A14 as a molecular biomarker for OSCCs (Paper III)

The S100A14 gene was selected for further in-depth analysis based on previous findings by our group and others, associating S100A14 with OSCC and esophageal cancer, being regulated by p53 (76, 77). Increased or decreased expression of S100A14 have also been observed in different cancers (91). In **Paper III**, IHC was performed and analyzed in OSCC tissue samples from Sri Lanka showing normal and malignant changes (**Figure 2 Paper III**). In the adjacent normal mucosa, S100A14 protein was found to be strongly expressed in the

cell membrane of the epithelial cells. Some of the epithelial cells showed cytoplasmic staining, but none showed nuclear staining. To a large extent, the dysplastic lesions showed the same expression as adjacent normal tissues, but more heterogeneously in the strength of expression. Across the OSCC tissue samples, a variable degree of staining was seen. The S100A14 protein expression was found to be clearly down-regulated and to some extent lost in the OSCC tissues, particularly in the invading tumor islands. The sub-cellular S100A14 expression pattern was also to some extent seen to change from plasma membrane to the cytoplasm. Loss and change in subcellular localization of the S100A14 expression from normal to cancerous tissue suggests a possible role for S100A14 in oral tumorigenesis, supporting our earlier findings (92). Further, a study of genetic variance in *S100A14* was performed on the two SNPs 461G>A and 1545A>T (**Table 3 Paper III**). Interestingly, the OSF samples from India were dominated by the heterozygotic genotype in the case of SNP 461G>A, with very few samples harboring the homozygotic alleles. The European (Norway, Sweden, UK) samples harbored the GG (38%) and GA (46.5%) genotype in a larger proportion than the AA genotype (15.5%). Interestingly, none of the Sudan samples harbored the AA genotype, only the GA or GG alleles, were distributed evenly. Chen *et al.* (77) demonstrated that patients bearing the 461A allele have a higher risk of developing esophageal cancer, particularly smokers bearing the AA genotype, and that this allele is associated with reduced expression of S100A14. Our results show that there is a larger proportion of patients from Asia (India and Sri Lanka) and Europe (Norway, Sweden, UK) harboring the A allele in total (89% and 62%, respectively) than in the Sudan (47%). This might be related to the different tobacco habits practiced in these regions. Our results might indicate that there is also a high risk of developing OSCC in patients with OSF harboring the 461A genotype, as 94% of the OSF cases were shown to be bearers of the A allele. This observation is of importance and warrants further in-depth studies with focus on the study of a large number of OSF cases. For the 1545A>T SNP, the allele frequency for OSFs also deviated from the European and African samples, with only one sample representing the AA genotype.

5. Conclusions

Microarray technology is a powerful tool to identify distinct patterns of gene expression and genome-wide DNA copy number variations in OSCCs from different populations.

Development of OSCCs follows a rather uniform pattern of biological pathways regardless of demographic differences related to ethnicity or risk factors. This observation is based on the findings that there are several common genes found to be involved in OSCCs from different populations.

Alterations in biological pathways involving several genes like *caveolins, matrix metalloproteinases, extracellular matrix, proteases, oncogenes and S100A gene family members are commonly found in OSCCs from different populations.*

The majority of the genetic changes observed seem to occur early in OSCC development also at premalignant stages. In this study changes in the S100 gene family have emerged as playing a key role in carcinogenesis.

Differential expression of genes may have dual roles in different OSCC samples, and between different types of cancers.

Down-regulation of the expression of the S100A14 is frequent in OSCC. This observation suggests that this protein might be used as a molecular biomarker in OSCC and OSF.

Findings of differential distribution of the A allele of the S100A14 gene in the SNP 461G>A among OSCCs and OSF from Asia, Africa and Europe might be related to tobacco habits and oral cancer susceptibility. This gene locus may be a particular target for the transition of OSF into OSCC.

6. Future perspectives

Proteomic analysis of the OSCC samples used in this study for the cDNA microarrays and array-CGH will compliment genomic studies reported here.

Performing stable isotope labeling with amino acids in cell culture (SILAC) using OSCC cell lines over-expressing S100A14, searching for differential protein expression.

Further studies on the functional role of S100A14 in OSCC to elucidate its involvement in cell cycle regulation and oral carcinogenesis.

Further studies on the association between the genetic variants of *S100A14* and OSCC/OSF, harvesting samples from Asian patients diagnosed with oral premalignant disorders.

Investigate the role of S100A14 as a molecular biomarker in field survey studies involving chronic smokeless tobacco users and patients diagnosed with oral mucosal changes using non-invasive methodologies.

Further studies on the possible involvement of the S100 gene family members *S100A7*, *S100A8* and *S100A9* in oral carcinogenesis.

7. References

1. International Agency for Research on Cancer: IARC Screening Group. http://screening.iarc.fr
2. WARNAKULASURIYA S. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol 2009; 45(4-5): 309-16.
3. MIGNOGNA MD, FEDELE S, RUSSO LL. The World Cancer Report and the burden of oral cancer. Eur J Cancer Prev 2003; 13: 139-142.
4. WARNAKULASURIYA S, JOHNSON NW, VAN DER WAAL I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. J Oral Pathol Med 2007; 36(10): 575-80.
5. FERLAY J, SHIN H, BRAY F, FORMAN D, MATHERS C, PARKIN D. Globocan 2008: Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. 2010; International Agency for Research on Cancer, Lyon.
6. LEE CH, KO AM, WARNAKULASURIYA S, YIN BL, ZAIN RB, IBRAHIM SO, LIU ZW, LI WH, ZHANG SS, UTOMO B, RAJAPAKSE PS, WARUSAVITHANA SA, RAZAK IA, ABDULLAH N, SHRESTHA P, KWAN AL, SHIEH TY, CHEN MK, KO YC. Inter-country prevalences and practices of betel-quid use in south, south east and eastern asia regions and associated oral preneoplastic disorders: An international collaborative study by asian betel-quid consortium of south and east asia. Int J Cancer 2010.
7. STEWART B, KLEIHUES P. World Cancer Report. 2003; International Agency for Research on Cancer, Lyon.
8. 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(4): 251-262.
9. IARC. Betel-quid and areca nut chewing and some areca-nut derived nitrosamines. IARC Monographs 2004; International Agency on Research on Cancer, Lyon.
10. CNATTINGIUS S, GALANTI R, GRAFSTRÖM R, HERGENS M-P, LAMBE M, NYRÉN O, PERSHAGEN G, WICKHOLM S. Hälsorisker med svenskt snus. R2005: 15, Statens folkhälsoinstitut.
11. CONWAY D, PETTICREW M, MARLBOROUGH H, BERTHILLER J, HASHIBE M, MACPHERSON L. Socioeconomic inequalitities and oral cancer risk: A systematic rewiev and meta-analysis of case-control studies. Int J Cancer 2008; 122: 2811-819.
12. LEWIN F, NORELL SE, JOHANSSON H, GUSTAVSSON P, WENNERBERG J, BIÖRKLUND A, RUTQVIST LE. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck. Cancer 1998; 82(7): 1367-1375.
13. RODRIGUES VC, MOSS SM, TUOMAINEN H. Oral cancer in the UK: to screen or not to screen. Oral Oncology 1998; 34(6): 454-465.
14. ROOSAAR A, JOHANSSON ALV, SANDBORGH-ENGLUND G, AXÉLL T, NYRÉN O. Cancer and mortality among users and nonusers of snus. Int J Cancer 2008; 123: 168-73.
15. SCHILDT E-B, ERIKSSON M, HARDELL L, MAGNUSON A. Oral snuff, smoking habits and alcohol consumption in relation to oral cancer in a Swedish case-control study. Int J Cancer 1998; 77(3): 341-346.
16. SCULLY C, BAGAN J. Oral squamous cell carcinoma overview. Oral Oncol 2009; 45(4-5): 301-308.

17. WARNAKULASURIYA S. Causes of oral cancer - an appraisal of controversies. Br Dent J 2009; 207(10): 471-475.
18. PINDBORG JJ, SIRSAT SM. Oral submucous fibrosis. Oral Surg Oral Med Oral Pathol 1966; 22(6): 764-79.
19. AVON SL. Oral mucosal lesions associated with use of quid. J Can Dent Assoc 2004; 70(4): 244-248.
20. MURTI PR, BHONSLE RB, PINDBORG JJ, DAFTARY DK, GUPTA PC, MEHTA FS. Malignant transformation rate in oral submucous fibrosis over a 17-year period. Community Dent Oral Epidemiol 1985; 13(6): 340-1.
21. NEVILLE BW, DAY TA. Oral cancer and precancerous lesions. CA Cancer J Clin 2002; 52(4): 195-215.
22. VAN DER WAAL I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. Oral Oncol 2003; 45(4-5): 317-323.
23. PETTI S. Pooled estimate of world leukoplakia prevalence: a systematic review. Oral Oncol 2003; 39(8): 770-80.
24. ROOPASHREE MR, GONDHALEKAR RV, SHASHIKANTH MC, GEORGE J, THIPPESWAMY SH, SHUKLA A. Pathogenesis of oral lichen planus-a review. J Oral Pathol Med 2010; 39(10): 729-34.
25. SCHLOSSER BJ. Lichen planus and lichenoid reactions of the oral mucosa. Dermatol Ther 2010; 23(3): 251-67.
26. SCULLY C, CARROZZO M. Oral mucosal disease: Lichen planus. Br J Oral Maxillofac Surg 2008; 46(1): 15-21.
27. HECHT SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev Cancer 2003; 3(10): 733-44.
28. JOHNSON NW, WARNAKULASURIYA S, GUPTA PC, DIMBA E, CHINDIA M, OTOH EC, SANKARANARAYANAN R, CALIFANO J, KOWALSKI L. Global oral health inequalities in incidence and outcomes for oral cancer: causes and solutions. Adv Dent Res 2011; 23(2): 237-46.
29. WARNAKULASURIYA S, SUTHERLAND G, SCULLY C. Tobacco, oral cancer, and treatment of dependence. Oral Oncol 2005; 41(3): 244-60.
30. RODU B, JANSSON C. Smokeless tobacco and oral cancer: a review of the risks and determinants. Crit Rev Oral Biol Med 2004; 15(5): 252-63.
31. BOFFETTA P, HECHT S, GRAY N, GUPTA P, STRAIF K. Smokeless tobacco and cancer. Lancet Oncol 2008; 9(7): 667-75.
32. LUO J, YE W, ZENDEHDEL K, ADAMI J, ADAMI H-O, BOFFETTA P, NYRÉN O. Oral use of Swedish moist snuff (snus) and risk for cancer of the mouth, lung, and pancreas in male construction workers: a retrospective cohort study. Lancet 2007; 369(9578): 2015-2020.
33. SCHILDT E-B, ERIKSSON M, HARDELL L, MAGNUSON A. Oral snuff, smoking habits and alcohol consumption in relation to oral cancer in a Swedish case-control study. Int J Cancer 1998; 77(3): 341-346.
34. BOFFETTA P, AAGNES B, WEIDERPASS E, ANDERSEN A. Smokeless tobacco use and risk of cancer of the pancreas and other organs. Int J Cancer 2005; 114(6): 992-5.
35. AHMED HG, MAHGOOB RM. Impact of Toombak dipping in the etiology of oral cancer: gender-exclusive hazard in the Sudan. J Cancer Res Ther 2007; 3(2): 127-30.
36. IDRIS AM, IBRAHIM SO, VASSTRAND EN, JOHANNESSEN AC, LILLEHAUG JR, MAGNUSSON B, WALLSTROM M, HIRSCH JM, NILSEN R. The Swedish snus and the Sudanese toombak: are they different? Oral Oncol 1998; 34(6): 558-66.

37. STANFILL SB, CONNOLLY GN, ZHANG L, JIA LT, HENNINGFIELD JE, RICHTER P, LAWLER TS, AYO-YUSUF OA, ASHLEY DL, WATSON CH. Global surveillance of oral tobacco products: total nicotine, unionised nicotine and tobacco-specific N-nitrosamines. Tob Control 2011; 20(3): e2.
38. COGLIANO V, STRAIF K, BAAN R, GROSSE Y, SECRETAN B, EL GHISSASSI F. Smokeless tobacco and tobacco-related nitrosamines. Lancet Oncol 2004; 5(12): 708.
39. IDRIS AM, AHMED HM, MALIK MO. Toombak dipping and cancer of the oral cavity in the Sudan: a case-control study. Int J Cancer 1995; 63(4): 477-80.
40. GUPTA PC, RAY CS. Smokeless tobacco and health in India and South Asia. Respirology 2003; 8(4): 419-31.
41. GUPTA PC, WARNAKULASURIYA S. Global epidemiology of areca nut usage. Addict Biol 2002; 7(1): 77-83.
42. VOGELSTEIN B, KINZLER KW. The multistep nature of cancer. Trends Genet 1993; 9(4): 138-41.
43. PELENGARIS S, KHAN M. The Molecular Biology of Cancer. Blackwell Publishing Ltd: Oxford, United Kingdom, 2006.
44. HANAHAN D, WEINBERG R. The Hallmarks of Cancer. Cell 2000; 100: 57-70.
45. 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(3): 256-263.
46. MCILWAIN CC, TOWNSEND DM, TEW KD. Glutathione S-transferase polymorphisms: cancer incidence and therapy. Oncogene 2006; 25(11): 1639-48.
47. AGRAWAL D, GUPTA S, AGARWAL D, GUPTA OP, AGARWAL M. Role of GSTM1 and GSTT1 polymorphism: susceptibility to oral submucous fibrosis in the North Indian population. Oncology 2010; 79(3-4): 181-6.
48. BUCH SC, NOTANI PN, BHISEY RA. Polymorphism at GSTM1, GSTM3 and GSTT1 gene loci and susceptibility to oral cancer in an Indian population. Carcinogenesis 2002; 23(5): 803-7.
49. CHEN MK, TSAI HT, CHUNG TT, SU SC, KAO TY, TSENG HC, LIU TH, CHIOU HL, YANG SF. Glutathione S-transferase P1 and alpha gene variants; role in susceptibility and tumor size development of oral cancer. Head Neck 2010; 32(8): 1079-87.
50. VOGELSTEIN B, KINZLER K. Cancer genes and the pathways they control. Nat Med 2004; 10(8): 789-99.
51. SCULLY C, FIELD JK, TANZAWA H. Genetic aberrations in oral or head and neck squamous cell carcinoma 2: chromosomal aberrations. Oral Oncol 2000; 36(4): 311-327.
52. TSANTOULIS PK, KASTRINAKIS NG, TOURVAS AD, LASKARIS G, GORGOULIS VG. Advances in the biology of oral cancer. Oral Oncol 2007; 43(6): 523-34.
53. CALIFANO J, VAN DER RIET P, WESTRA W, NAWROZ H, CLAYMAN G, PIANTADOSI S, CORIO R, LEE D, GREENBERG B, KOCH W, SIDRANSKY D. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer Res 1996; 56(11): 2488-92.
54. HA PK, CHANG SS, GLAZER CA, CALIFANO JA, SIDRANSKY D. Molecular techniques and genetic alterations in head and neck cancer. Oral Oncol 2009; 45(4): 335-339.

55. RAYBAUD-DIOGENE H, TETU B, MORENCY R, FORTIN A, MONTEIL RA. p53 overexpression in head and neck squamous cell carcinoma: review of the literature. Eur J Cancer B Oral Oncol 1996; 32B(3): 143-9.
56. GOH AM, COFFILL CR, LANE DP. The role of mutant p53 in human cancer. J Pathol 2011; 223(2): 116-26.
57. IBRAHIM SO, VASSTRAND EN, JOHANNESSEN AC, IDRIS AM, MAGNUSSON B, NILSEN R, LILLEHAUG JR. Mutations of the p53 gene in oral squamous-cell carcinomas from Sudanese dippers of nitrosamine-rich toombak and non-snuff-dippers from the Sudan and Scandinavia. Int J Cancer 1999; 81(4): 527-34.
58. WARNAKULASURIYA KA, RALHAN R. Clinical, pathological, cellular and molecular lesions caused by oral smokeless tobacco--a review. J Oral Pathol Med 2007; 36(2): 63-77.
59. CHOI S, MYERS JN. Molecular pathogenesis of oral squamous cell carcinoma: implications for therapy. J Dent Res 2008; 87(1): 14-32.
60. VOUSDEN KH, LU X. Live or let die: the cell's response to p53. Nat Rev Cancer 2002; 2(8): 594-604.
61. PITIYAGE G, TILAKARATNE WM, TAVASSOLI M, WARNAKULASURIYA S. Molecular markers in oral epithelial dysplasia: review. J Oral Pathol Med 2009; 38(10): 737-752.
62. SCHWARTZ JL. Biomarkers and molecular epidemiology and chemoprevention of oral carcinogenesis. Crit Rev Oral Biol Med 2000; 11(1): 92-122.
63. MARENHOLZ I, HEIZMANN CW, FRITZ G. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). Biochem Biophys Res Commun 2004; 322(4): 1111-1122.
64. SALAMA I, MALONE PS, MIHAIMEED F, JONES JL. A review of the S100 proteins in cancer. Eur J Surg Oncol 2008; 34(4): 357-364.
65. GENDLER SJ, COHEN EP, CRASTON A, DUHIG T, JOHNSTONE G, BARNES D. The locus of the polymorphic epithelial mucin (PEM) tumour antigen on chromosome 1q21 shows a high frequency of alteration in primary human breast tumours. Int J Cancer 1990; 45(3): 431-5.
66. WETERMAN MA, WILBRINK M, DIJKHUIZEN T, VAN DEN BERG E, GEURTS VAN KESSEL A. Fine mapping of the 1q21 breakpoint of the papillary renal cell carcinoma-associated (X;1) translocation. Hum Genet 1996; 98(1): 16-21.
67. MOORE BW. A soluble protein characteristic of the nervous system. Biochem Biophys Res Commun 1965; 19(6): 739-44.
68. DONATO R. Intracellular and extracellular roles of S100 proteins. Microscopy Research and Technique 2003; 60(6): 540-551.
69. SANTAMARIA-KISIEL L, RINTALA-DEMPSEY AC, SHAW GS. Calcium-dependent and -independent interactions of the S100 protein family. Biochem J 2006; 396(2): 201-14.
70. FRITZ G, BOTELHO HM, MOROZOVA-ROCHE LA, GOMES CM. Natural and amyloid self-assembly of S100 proteins: structural basis of functional diversity. FEBS J 2010; 277(22): 4578-90.
71. EMBERLEY ED, MURPHY LC, WATSON PH. S100 proteins and their influence on pro-survival pathways in cancer. Biochem Cell Biol 2004; 82(4): 508-15.
72. CHEN H, FERNIG DG, RUDLAND PS, SPARKS A, WILKINSON MC, BARRACLOUGH R. Binding to intracellular targets of the metastasis-inducing protein, S100A4 (p9Ka). Biochem Biophys Res Commun 2001; 286(5): 1212-7.
73. GRIGORIAN M, ANDRESEN S, TULCHINSKY E, KRIAJEVSKA M, CARLBERG C, KRUSE C, COHN M, AMBARTSUMIAN N, CHRISTENSEN A, SELIVANOVA

G, LUKANIDIN E. Tumor suppressor p53 protein is a new target for the metastasis-associated Mts1/S100A4 protein: functional consequences of their interaction. J Biol Chem 2001; 276(25): 22699-708.
74. FERNANDEZ-FERNANDEZ MR, RUTHERFORD TJ, FERSHT AR. Members of the S100 family bind p53 in two distinct ways. Protein Sci 2008; 17(10): 1663-70.
75. MUELLER A, SCHAFER BW, FERRARI S, WEIBEL M, MAKEK M, HOCHLI M, HEIZMANN CW. The calcium-binding protein S100A2 interacts with p53 and modulates its transcriptional activity. J Biol Chem 2005; 280(32): 29186-93.
76. SAPKOTA D, COSTEA D, BLØ M, BRULAND O, LORENS J, VASSTRAND EN, IBRAHIM S. S100A14 induces G1-arrest through nuclear accumulation of p53 and p53-dependent up-regulation of p21 in oral carcinoma derived cells. Oral Oncol 2011; In Press.
77. CHEN H, YU D, LUO A, TAN W, ZHANG C, ZHAO D, YANG M, LIU J, LIN D, LIU Z. Functional role of S100A14 genetic variants and their association with esophageal squamous cell carcinoma. Cancer Res 2009; 69(8): 3451-7.
78. SAPKOTA D, BRULAND O, BØE OE, BAKEER H, ELGINDI OAA, VASSTRAND EN, IBRAHIM SO. Expression profile of the S100 gene family members in oral squamous cell carcinomas. J Oral Pathol Med 2008; 37(10): 607-615.
79. JI J, ZHAO L, WANG X, ZHOU C, DING F, SU L, ZHANG C, MAO X, WU M, LIU Z. Differential expression of S100 gene family in human esophageal squamous cell carcinoma. J Cancer Res Clin Oncol 2004; 130(8): 480-486.
80. DYSVIK B, VASSTRAND EN, LØVLIE R, ELGINDI OAA, KROSS KW, AARSTAD JH, JOHANNESSEN AC, JONASSEN I, IBRAHIM SO. Gene expression profiles of head and neck carcinomas from Sudanese and Norwegian patients reveal common biological pathways regardless of race and lifestyle. Clin Cancer Res 2006; 12(4): 1109-20.
81. ROMAN E, MEZA-ZEPEDA L, KRESSE S, MYKLEBOST O, VASSTRAND E, IBRAHIM S. Chromosomal aberrations in head and neck squamous cell carcinomas in Norwegian and Sudanese populations by array comparative genomic hybridization. Oncol Rep 2008; 20: 825-43.
82. TSAI S-T, JIN Y-T, TSAI W-C, WANG S-T, LIN Y-C, CHANG M-T, WU L-W. S100A2, a potential marker for early recurrence in early-stage oral cancer. Oral Oncol 2005; 41(4): 349-357.
83. TSAI WC, TSAI ST, JIN YT, WU LW. Cyclooxygenase-2 is involved in S100A2-mediated tumor suppression in squamous cell carcinoma. Mol Cancer Res 2006; 4(8): 539-47.
84. YANG X, WEI KJ, ZHANG L, PAN HY, YE DX, ZHONG LP, ZHANG ZY. Decreased expression of S100A6 in oral squamous cell carcinoma. Oncol Rep 2010; 24(2): 479-88.
85. KESTING MR, SUDHOFF H, HASLER RJ, NIEBERLER M, PAUTKE C, WOLFF K-D, WAGENPFEIL S, AL-BENNA S, JACOBSEN F, STEINSTRAESSER L. Psoriasin (S100A7) up-regulation in oral squamous cell carcinoma and its relation to clinicopathologic features. Oral Oncol 2009; 45(8): 731-736.
86. TRIPATHI SC, MATTA A, KAUR J, GRIGULL J, CHAUHAN SS, THAKAR A, SHUKLA NK, DUGGAL R, DATTAGUPTA S, RALHAN R, SIU KW. Nuclear S100A7 is associated with poor prognosis in head and neck cancer. PLoS One 2010; 5(8): e11939.
87. WENGHOEFER M, PANTELIS A, NAJAFI T, DESCHNER J, ALLAM JP, NOVAK N, REICH R, MARTINI M, BERGE S, FISCHER HP, JEPSEN S, WINTER J. Gene expression of oncogenes, antimicrobial peptides, and cytokines in the

development of oral leukoplakia. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2010; 110(3): 351-6.
88. ZHOU G, XIE TX, ZHAO M, JASSER SA, YOUNES MN, SANO D, LIN J, KUPFERMAN ME, SANTILLAN AA, PATEL V, GUTKIND JS, EI-NAGGAR AK, EMBERLEY ED, WATSON PH, MATSUZAWA SI, REED JC, MYERS JN. Reciprocal negative regulation between S100A7/psoriasin and beta-catenin signaling plays an important role in tumor progression of squamous cell carcinoma of oral cavity. Oncogene 2008; 27(25): 3527-38.
89. LO WY, LAI CC, HUA CH, TSAI MH, HUANG SY, TSAI CH, TSAI FJ. S100A8 is identified as a biomarker of HPV18-infected oral squamous cell carcinomas by suppression subtraction hybridization, clinical proteomics analysis, and immunohistochemistry staining. J Proteome Res 2007; 6(6): 2143-51.
90. THIEL UJE, FELTENS R, ADRYAN B, GIERINGER R, BROCHHAUSEN C, SCHUON R, FILLIES T, GRUS F, MANN WJ, BRIEGER J. Analysis of differentially expressed proteins in oral squamous cell carcinoma by MALDI-TOF MS. J Oral Pathol Med 2011; 40(5): 369-379.
91. PIETAS A, SCHLUNS K, MARENHOLZ I, SCHAFER BW, HEIZMANN CW, PETERSEN I. Molecular cloning and characterization of the human S100A14 gene encoding a novel member of the S100 family. Genomics 2002; 79(4): 513-22.
92. SAPKOTA D, BRULAND O, COSTEA DE, HAUGEN H, VASSTRAND EN, IBRAHIM SO. S100A14 regulates the invasive potential of oral squamous cell carcinoma derived cell-lines in vitro by modulating expression of matrix metalloproteinases, MMP1 and MMP9. Eur J Cancer; 47(4): 600-610.
93. CHURCHILL GA. Fundamentals of experimental design for cDNA microarrays. Nat Genet 2002; 32 Suppl: 490-5.
94. MANNING AT, GARVIN JT, SHAHBAZI RI, MILLER N, MCNEILL RE, KERIN MJ. Molecular profiling techniques and bioinformatics in cancer research. Eur J Surg Oncol 2007; 33(3): 255-65.
95. NAZMUL-HOSSAIN AN, PATEL KJ, RHODUS NL, MOSER KL. Microarrays: applications in dental research. Oral Dis 2008; 14(1): 25-9.
96. MILLER MB, TANG YW. Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol Rev 2009; 22(4): 611-33.
97. CHOI P, CHEN C. Genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma. Cancer 2005; 104(6): 1113-28.
98. VIET CT, SCHMIDT BL. Understanding oral cancer in the genome era. Head Neck 2010; 32(9): 1246-68.
99. IBRAHIM SO, AARSAETHER N, HOLSVE MK, KROSS KW, HEIMDAL JH, AARSTAD JH, LIAVAAG PG, ELGINDI OAA, JOHANNESSEN AC, LILLEHAUG JR, VASSTRAND EN. Gene expression profile in oral squamous cell carcinomas and matching normal oral mucosal tissues from black Africans and white Caucasians: the case of the Sudan vs. Norway. Oral Oncol 2003; 39(1): 37-48.
100. BELBIN TJ, BARBER I, SOCCI N, WENIG B, SMITH R, PRYSTOWSKY MB, CHILDS G. Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays. Cancer Res 2004; 62: 1184-1190.
101. CHUNG CH, PARKER JS, KARACA G, WU J, FUNKHOUSER WK, MOORE D, BUTTERFOSS D, XIANG D, ZANATION A, YIN X, SHOCKLEY WW, WEISSLER MC, DRESSLER LG, SHORES CG, YARBROUGH WG, PEROU CM. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 2004; 5(5): 489-500.

102. HA PK, BENOIT NE, YOCHEM R, SCIUBBA J, ZAHURAK M, SIDRANSKY D, PEVSNER J, WESTRA WH, CALIFANO J. A transcriptional progression model for head and neck cancer. Clin Cancer Res 2003; 9(8): 3058-64.
103. ZIOBER AF, PATEL KR, ALAWI F, GIMOTTY P, WEBER RS, FELDMAN MM, CHALIAN AA, WEINSTEIN GS, HUNT J, ZIOBER BL. Identification of a gene signature for rapid screening of oral squamous cell carcinoma. Clin Cancer Res 2006; 12(20 Pt 1): 5960-71.
104. CHEN C, MENDEZ E, HOUCK J, FAN W, LOHAVANICHBUTR P, DOODY D, YUEH B, FUTRAN ND, UPTON M, FARWELL DG, SCHWARTZ SM, ZHAO LP. Gene Expression Profiling Identifies Genes Predictive of Oral Squamous Cell Carcinoma. Cancer Epidemiol Biomarkers Prev 2008; 17(8): 2152-2162.
105. KONDOH N, OHKURA S, ARAI M, HADA A, ISHIKAWA T, YAMAZAKI Y, SHINDOH M, TAKAHASHI M, KITAGAWA Y, MATSUBARA O, YAMAMOTO M. Gene expression signatures that can discriminate oral leukoplakia subtypes and squamous cell carcinoma. Oral Oncol 2007; 43(5): 455-62.
106. BELBIN TJ, SINGH B, SMITH RV, SOCCI ND, WREESMANN VB, SANCHEZ-CARBAYO M, MASTERSON J, PATEL S, CORDON-CARDO C, PRYSTOWSKY MB, CHILDS G. Molecular profiling of tumor progression in head and neck cancer. Arch Otolaryngol Head Neck Surg 2005; 131(1): 10-8.
107. CROMER A, CARLES A, MILLON R, GANGULI G, CHALMEL F, LEMAIRE F, YUONG J, DEMBÈLÈ D, THIBAULT C, MULLER D, POCH O, ABECASSIS J, WASYLYK B. Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. Oncogene 2004; 23: 2484-498.
108. NGUYEN ST, HASEGAWA S, TSUDA H, TOMIOKA H, USHIJIMA M, NODA M, OMURA K, MIKI Y. Identification of a predictive gene expression signature of cervical lymph node metastasis in oral squamous cell carcinoma. Cancer Sci 2007; 98(5): 740-6.
109. 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.
110. ROEPMAN P, WESSELS LF, KETTELARIJ N, KEMMEREN P, MILES AJ, LIJNZAAD P, TILANUS MG, KOOLE R, HORDIJK GJ, VAN DER VLIET PC, REINDERS MJ, SLOOTWEG PJ, HOLSTEGE FC. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. Nat Genet 2005; 37(2): 182-6.
111. SCHMALBACH CE, CHEPEHA DB, GIORDANO TJ, RUBIN MA, TEKNOS TN, BRADFORD CR, WOLF GT, KUICK R, MISEK DE, TRASK DK, HANASH S. Molecular profiling and the identification of genes associated with metastatic oral cavity/pharynx squamous cell carcinoma. Arch Otolaryngol Head Neck Surg 2004; 130(3): 295-302.
112. WARNER GC, REIS PP, JURISICA I, SHILPI MS, MACMILLAN AC, MAKITIE AA, GRÉNMAN R, REID N, SUKHAI M, FREEMAN J, GULLANE P, IRISH J, KAMEL-REID S. Molecular classification of oral cancer by cDNA microarrays identifies overexpressed genes correlated with nodal metastasis. Int J Cancer 2004; 110(6): 857-868.
113. ZHOU S, KASSAUEI K, CUTLER DJ, KENNEDY GC, SIDRANSKY D, MAITRA A, CALIFANO J. An oligonucleotide microarray for high-throughput sequencing of the mitochondrial genome. J Mol Diagn 2006; 8(4): 476-82.

114. CHEONG SC, CHANDRAMOULI GVR, SALEH A, ZAIN RB, LAU SH, SIVAKUMAREN S, PATHMANATHAN R, PRIME SS, TEO SH, PATEL V, GUTKIND JS. Gene expression in human oral squamous cell carcinoma is influenced by risk factor exposure. Oral Oncol 2009; 45(8): 712-719.
115. SHINAWI M, CHEUNG SW. The array CGH and its clinical applications. Drug Discov Today 2008; 13(17-18): 760-70.
116. SNIJDERS AM, PINKEL D, ALBERTSON DG. Current status and future prospects of array-based comparative genomic hybridisation. Brief Funct Genomic Proteomic 2003; 2(1): 37-45.
117. BEAUDET AL, BELMONT JW. Array-based DNA diagnostics: let the revolution begin. Annu Rev Med 2008; 59: 113-29.
118. PINKEL D, ALBERTSON DG. Comparative genomic hybridization. Annu Rev Genomics Hum Genet 2005; 6: 331-54.
119. RAPAPORT F, LESLIE C. Determining frequent patterns of copy number alterations in cancer. PLoS One 2010; 5(8): e12028.
120. REDON R, ISHIKAWA S, FITCH KR, FEUK L, PERRY GH, ANDREWS TD, FIEGLER H, SHAPERO MH, CARSON AR, CHEN W, CHO EK, DALLAIRE S, FREEMAN JL, GONZALEZ JR, GRATACOS M, HUANG J, KALAITZOPOULOS D, KOMURA D, MACDONALD JR, MARSHALL CR, MEI R, MONTGOMERY L, NISHIMURA K, OKAMURA K, SHEN F, SOMERVILLE MJ, TCHINDA J, VALSESIA A, WOODWARK C, YANG F, ZHANG J, ZERJAL T, ARMENGOL L, CONRAD DF, ESTIVILL X, TYLER-SMITH C, CARTER NP, ABURATANI H, LEE C, JONES KW, SCHERER SW, HURLES ME. Global variation in copy number in the human genome. Nature 2006; 444(7118): 444-54.
121. BALDWIN C, GARNIS C, ZHANG L, ROSIN MP, LAM WL. Multiple microalterations detected at high frequency in oral cancer. Cancer Res 2005; 65(17): 7561-7.
122. HERMSEN M, JOENJE H, ARWERT F, BRAAKHUIS B, BAAK J, WESTERVELD A, SLATER R. Assessment of chromosomal gains and losses in oral squamous cell carcinoma by comparative genomic hybridisation. Oral Oncol 1997; 33: 414-18.
123. NAKAMURA E, KOZAKI K, TSUDA H, SUZUKI E, PIMKHAOKHAM A, YAMAMOTO G, IRIE T, TACHIKAWA T, AMAGASA T, INAZAWA J, IMOTO I. Frequent silencing of a putative tumor suppressor gene melatonin receptor 1 A (MTNR1A) in oral squamous-cell carcinoma. Cancer Sci 2008; 99(7): 1390-400.
124. PATMORE H, CAWKWELL L, STAFFORD N, GREENMAN J. Unraveling the Chromosomal Aberrations of Head and Neck Squamous Cell Carcinoma: A Review. Ann Surg Oncol 2005; 12(10): 831-842.
125. SUZUKI E, IMOTO I, PIMKHAOKHAM A, NAKAGAWA T, KAMATA N, KOZAKI KI, AMAGASA T, INAZAWA J. PRTFDC1, a possible tumor-suppressor gene, is frequently silenced in oral squamous-cell carcinomas by aberrant promoter hypermethylation. Oncogene 2007; 26(57): 7921-32.
126. WOLFF E, GIROD S, LIEHR T, VORDERWULBECKE U, RIES J, STEININGER H, GEBHART E. Oral squamous cell carcinomas are characterized by a rather uniform pattern of genomic imbalances detected by comparative genomic hybridisation. Oral Oncol 1998; 34: 186-90.
127. MEZA-ZEPEDA LA, KRESSE SH, BARRAGAN-POLANIA AH, BJERKEHAGEN B, OHNSTAD HO, NAMLOS HM, WANG J, KRISTIANSEN BE, MYKLEBOST O. Array comparative genomic hybridization reveals distinct DNA copy number differences between gastrointestinal stromal tumors and leiomyosarcomas. Cancer Res 2006; 66(18): 8984-93.

128. OKAFUJI M, ITA M, OGA A, HAYATSU Y, MATSUO A, SHINZATO Y, SHINOZAKI F, SASAKI K. The relationship of genetic aberrations detected by comparative genomic hybridization to DNA ploidy and tumor size in human oral squamous cell carcinomas. J Oral Pathol Med 2000; 29(5): 226-31.
129. WEBER R, SCHEER M, BORN I, JOOS S, COBBERS J, HOFELE C, REIFENBERGER G, ZÕLLER J, LICHTER P. Recurrent chromosomal imbalances detected in biopsy material from oral premalignant and malignant lesions by combined tissue microdissection, universal DNA amplification and comparative genomic hybridization. Am J Pathol 1998; 153: 295-303.
130. GARNIS C, CHARI R, BUYS TP, ZHANG L, NG RT, ROSIN MP, LAM WL. Genomic imbalances in precancerous tissues signal oral cancer risk. Mol Cancer 2009; 8: 50.
131. LIN SC, CHEN YJ, KAO SY, HSU MT, LIN CH, YANG SC, LIU TY, CHANG KW. Chromosomal changes in betel-associated oral squamous cell carcinomas and their relationship to clinical parameters. Oral Oncol 2002; 38(3): 266-73.
132. CHATTOPADHYAY I, SINGH A, PHUKAN R, PURKAYASTHA J, KATAKI A, MAHANTA J, SAXENA S, KAPUR S. Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma exposed to tobacco and betel quid from high-risk area in India. Mutat Res 2010; 696(2): 130-8.
133. Applied Biosystems: Real-Time PCR. www.appliedbiosystems.com
134. VANGUILDER HD, VRANA KE, FREEMAN WM. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 2008; 44(5): 619-26.
135. JENKINS GJS, WILLIAMS GL, BEYNON J, YE Z, BAXTER JN, PARRY JM. Restriction enzymes in the analysis of genetic alterations responsible for cancer progression. Br J Surg 2002; 89(1): 8-20.
136. CARVALHO AL, PINTOS J, SCHLECHT NF, OLIVEIRA BV, FAVA AS, CURADO MP, KOWALSKI LP, FRANCO EL. Predictive factors for diagnosis of advanced-stage squamous cell carcinoma of the head and neck. Arch Otolaryngol Head Neck Surg 2002; 128(3): 313-8.
137. BRANDTZAEG P. The increasing power of immunohistochemistry and immunocytochemistry. J Immunol Methods 1998; 216(1-2): 49-67.
138. KUMAR G, RUDBECK L. Immunohistochemichal Staining Methods. Fifth edition. DAKO: www.dako.com
139. EFRON B, TIBSHIRANI R, GOSS V, CHU G. Microarrays and their use in a comparative experiment. Technical Report, 2000; Division of Biostatistics, Stanford University.
140. YANG Y, DUDOIT S, LUU P, LIN D, PENG V, NGAI J, SPEED T. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation Nucleic Acids Res 2002; 30(4): e15.
141. COHEN AW, HNASKO R, SCHUBERT W, LISANTI MP. Role of caveolae and caveolins in health and disease. Physiol Rev 2004; 84(4): 1341-79.
142. ENGELMAN JA, ZHANG X, GALBIATI F, VOLONTÈ D, SOTGIA F, PESTELL RG, MINETTI C, SCHERER PE, OKAMOTO T, LISANTI MP. Molecular Genetics of the Caveolin Gene Family: Implications for Human Cancers, Diabetes, Alzheimer Disease, and Muscular Dystrophy. Am J Hum Genet 1998; 63: 1578-587.
143. ANDO T, ISHIGURO H, KIMURA M, MITSUI A, MORI Y, SUGITO N, TOMODA K, MORI R, HARADA K, KATADA T, OGAWA R, FUJII Y, KUWABARA Y. The overexpression of caveolin-1 and caveolin-2 correlates with a poor prognosis and tumor progression in esophageal squamous cell carcinoma. Oncol Rep 2007; 18(3): 601-9.

144. HAN SE, PARK KH, LEE G, HUH YJ, MIN BM. Mutation and aberrant expression of Caveolin-1 in human oral squamous cell carcinomas and oral cancer cell lines. Int J Oncol 2004; 24(2): 435-40.
145. NAKATANI K, WADA T, NAKAMURA M, UZAWA K, TANZAWA H, FUJITA S. Expression of caveolin-1 and its correlation with cisplatin sensitivity in oral squamous cell carcinoma. J Cancer Res Clin Oncol 2005; 131(7): 445-452.
146. XUE J, CHEN H, DIAO L, CHEN X, XIA D. Expression of caveolin-1 in tongue squamous cell carcinoma by quantum dots. Eur J Histochem 2010; 54(2): e20.
147. IBRAHIM S, MIRON T, KROHN M, AMARATUNGA A, WARNAKULASURIYA S, VASSTRAND E. Tumour-associated Proteins in Oral Squamous Cell Carcinomas by Proteomics. Cancer Genomics Proteomics 2005; 2: 353-364.
148. LYONS AJ, JONES J. Cell adhesion molecules, the extracellular matrix and oral squamous carcinoma. Int J Oral Maxillofac Surg 2007; 36(8): 671-9.
149. THOMAS GT, LEWIS MP, SPEIGHT PM. Matrix metalloproteinases and oral cancer. Oral Oncol 1999; 35(3): 227-233.
150. YEN C-Y, CHEN C-H, CHANG C-H, TSENG H-F, LIU S-Y, CHUANG L-Y, WEN C-H, CHANG H-W. Matrix metalloproteinases (MMP) 1 and MMP10 but not MMP12 are potential oral cancer markers. Biomarkers 2009; 14(4): 244-249.
151. YE H, YU T, TEMAM S, ZIOBER B, WANG J, SCHWARTZ J, MAO L, WONG D, ZHOU X. Transcriptomic dissection of tongue squamous cell carcinoma. BMC Genomics 2008; 9(1): 69.
152. TSAI W, TSAI S, KO J, JIN Y, LI C, HUANG W, YOUNG K, LAI M, LIU H, WU L. The mRNA profile of genes in betel quid chewing oral cancer patients. Oral Oncol 2004; 40(4): 418-426.
153. DE VICENTE JC, FRESNO MF, VILLALAIN L, VEGA JA, HERNANDEZ VALLEJO G. Expression and clinical significance of matrix metalloproteinase-2 and matrix metalloproteinase-9 in oral squamous cell carcinoma. Oral Oncol 2005; 41(3): 283-93.
154. 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(9): 1902-910.
155. CHIU CJ, CHANG ML, CHIANG CP, HAHN LJ, HSIEH LL, CHEN CJ. Interaction of collagen-related genes and susceptibility to betel quid-induced oral submucous fibrosis. Cancer Epidemiol Biomarkers Prev 2002; 11(7): 646-53.
156. TAI IT, TANG MJ. SPARC in cancer biology: Its role in cancer progression and potential for therapy. Drug Resistance Updates 2008; 11(6): 231-246.
157. CHOI P, JORDAN CD, MENDEZ E, HOUCK J, YUEH B, FARWELL DG, FUTRAN N, CHEN C. Examination of oral cancer biomarkers by tissue microarray analysis. Arch Otolaryngol Head Neck Surg 2008; 134(5): 539-46.
158. ZHANG J, ZHANG Q, LIU XK, LI CQ, GUO ZM. Expression and clinical significance of SPARC in clinical stage II tongue squamous cell carcinoma. Ai Zheng 2009; 28(1): 68-71.
159. CHIQUET-EHRISMANN R, CHIQUET M. Tenascins: regulation and putative functions during pathological stress. J Pathol 2003; 200(4): 488-99.
160. OREND G. Potential oncogenic action of tenascin-C in tumorigenesis. Int J Biochem Cell Biol 2005; 37(5): 1066-83.
161. KALEMBEYI I, INADA H, NISHIURA R, IMANAKA-YOSHIDA K, SAKAKURA T, YOSHIDA T. Tenascin-C upregulates matrix metalloproteinase-9 in breast cancer cells: direct and synergistic effects with transforming growth factor beta1. Int J Cancer 2003; 105(1): 53-60.

162. DUFFY MJ, DUGGAN C. The urokinase plasminogen activator system: a rich source of tumour markers for the individualised management of patients with cancer. Clin Biochem 2004; 37(7): 541-548.
163. NOZAKI S, ENDO Y, NAKAHARA H, YOSHIZAWA K, HASHIBA Y, KAWASHIRI S, TANAKA A, NAKAGAWA K, MATSOUKA Y, KOGO M, YAMOMOTO E. Inhibition of invasion and metastasis in oral cancer by targeting urokinase-type plasminogen activation receptor. Oral Oncol 2005; 41: 971-77.
164. KATO Y, NAGASHIMA Y, BABA Y, KAWANO T, FURUKAWA M, KUBOTA A, YANOMA S, IMAGAWA-ISHIGURO Y, SATAKE K, TAGUCHI T, HATA R, MOCHIMATSU I, AOKI I, KAMEDA Y, INAYAMA Y, TSUKUDA M. Expression of SPARC in tongue carcinoma of stage II is associated with poor prognosis: An immunohistochemical study of 86 cases. Int J Mol Med 2005; 16: 263-68.
165. HORREVOETS A, FONTIJN R, VAN ZONNENVELD A, DE VRIES C, TEN CATE J, PANNKOEK H. Vascular Endothelial Genes That are Responsive to Tumor Necrosis Factor-alfa In Vitro Are Expressed in Atherosclerotic Lesions, Including Inhibitor of Apoptosis Protein-1, Stannin and Two Novel Genes. Blood 1999; 93(10): 3418-3431.
166. VANHOLLEBEKE B, PAYS E. The function of apolipoproteins L. Cell Mol Life Sci 2006; 63(17): 1937-44.
167. JOHANNESON B, MCDONNELL SK, KARYADI DM, QUIGNON P, MCINTOSH L, RISKA SM, FITZGERALD LM, JOHNSON G, DEUTSCH K, WILLIAMS G, TILLMANS LS, STANFORD JL, SCHAID DJ, THIBODEAU SN, OSTRANDER EA. Family-based association analysis of 42 hereditary prostate cancer families identifies the Apolipoprotein L3 region on chromosome 22q12 as a risk locus. Hum Mol Genet 2010; 19(19): 3852-62.
168. NICHOLSON TB, CHEN T, RICHARD S. The physiological and pathophysiological role of PRMT1-mediated protein arginine methylation. Pharmacol Res 2009; 60(6): 466-74.
169. JENUWEIN T, ALLIS CD. Translating the histone code. Science 2001; 293(5532): 1074-80.
170. LUND AH, VAN LOHUIZEN M. Epigenetics and cancer. Genes Dev 2004; 18(19): 2315-35.
171. AN W, KIM J, ROEDER RG. Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. Cell 2004; 117(6): 735-48.
172. MENENDEZ JA, LUPU R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. Nat Rev Cancer 2007; 7(10): 763-777.
173. MASHIMA T, SEIMIYA H, TSURUO T. De novo fatty-acid synthesis and related pathways as molecular targets for cancer therapy. Br J Cancer 2009; 100(9): 1369-72.
174. MILDE-LANGOSCH K. The Fos family of transcription factors and their role in tumourigenesis. Eur J Cancer 2005; 41(16): 2449-61.
175. JIN SP, KIM JH, KIM MA, YANG HK, LEE HE, LEE HS, KIM WH. Prognostic significance of loss of c-fos protein in gastric carcinoma. Pathol Oncol Res 2007; 13(4): 284-9.
176. MIKULA M, GOTZMANN J, FISCHER AN, WOLSCHEK MF, THALLINGER C, SCHULTE-HERMANN R, BEUG H, MIKULITS W. The proto-oncoprotein c-Fos negatively regulates hepatocellular tumorigenesis. Oncogene 2003; 22(43): 6725-38.
177. WRIGHT NT, CANNON BR, ZIMMER DB, WEBER DJ. S100A1: Structure, Function, and Therapeutic Potential. Curr Chem Biol 2009; 3(2): 138-145.

178. 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.
179. SUZUKI F, ORIDATE N, HOMMA A, NAKAMARU Y, NAGAHASHI T, YAGI K, YAMAGUCHI S, FURUTA Y, FUKUDA S. S100A2 expression as a predictive marker for late cervical metastasis in stage I and II invasive squamous cell carcinoma of the oral cavity. Oncol Rep 2005; 14(6): 1493-8.
180. LO JF, YU CC, CHIOU SH, HUANG CY, JAN CI, LIN SC, LIU CJ, HU WY, YU YH. The epithelial-mesenchymal transition mediator S100A4 maintains cancer-initiating cells in head and neck cancers. Cancer Res 2011; 71(5): 1912-23.
181. MORIYAMA-KITA M, ENDO Y, YONEMURA Y, HEIZMANN CW, SCHÄFER BW, SASAKI T, YAMAMOTO E. Correlation of S100A4 expression with invasion and metastasis in oral squamous cell carcinoma. Oral Oncol 2004; 40(5): 496-500.
182. CHIBA I, MUTHUMALA M, YAMAZAKI Y, UZ ZAMAN A, IIZUKA T, AMEMIYA A, SHIBATA T, KASHIWAZAKI H, SUGIURA C, FUKUDA H. Characteristics of mutations in the p53 gene of oral squamous-cell carcinomas associated with betel-quid chewing in Sri Lanka. Int J Cancer 1998; 77(6): 839-42.
183. RALHAN R, AGARWAL S, NATH N, MATHUR M, WASYLYK B, SRIVASTAVA A. Correlation between p53 gene mutations and circulating antibodies in betel- and tobacco-consuming North Indian population. Oral Oncol 2001; 37(3): 243-50.
184. CHEN IH, CHANG JT, LIAO CT, WANG HM, HSIEH LL, CHENG AJ. Prognostic significance of EGFR and Her-2 in oral cavity cancer in betel quid prevalent area cancer prognosis. Br J Cancer 2003; 89(4): 681-6.
185. MOON C, CHAE YK, LEE J. Targeting epidermal growth factor receptor in head and neck cancer: lessons learned from cetuximab. Exp Biol Med (Maywood) 2010; 235(8): 907-20.
186. LAI JP, OSEINI AM, MOSER CD, YU C, ELSAWA SF, HU C, NAKAMURA I, HAN T, ADERCA I, ISOMOTO H, GARRITY-PARK MM, SHIRE AM, LI J, SANDERSON SO, ADJEI AA, FERNANDEZ-ZAPICO ME, ROBERTS LR. The oncogenic effect of sulfatase 2 in human hepatocellular carcinoma is mediated in part by glypican 3-dependent Wnt activation. Hepatology 2010; 52(5): 1680-9.
187. NADANAKA S, KITAGAWA H. Heparan Sulphate Biosynthesis and Disease. J Biochem 2008; 144(1): 7-14.