Universidade de Aveiro Departamento de Química

2014

Elsa Maltez

Proteoma Salivar em pacientes com Cancro da Cabeça e do Pescoço

Elsa Maltez

Proteoma Salivar em pacientes com Cancro da Cabeça e do Pescoço

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção de grau de Mestre em Bioquímica, ramo de Bioquímica Clínica, realizada sob a orientação científica do Professor Doutor Francisco Manuel Lemos Amado, professor associado do Departamento de Quimica e director da Escola Superior de Saúde da Universidade de Aveiro e do Doutor Hugo Manuel Lopes Sousa, Grupo de Oncologia Molecular, Instituto Português de Oncologia Porto FG EPE.

Ao Ricardo e ao Eurico,

O júri

Presidente

Professor Doutor Pedro Miguel Dimas Neves Domingues

Professor Auxiliar com agregação do Departamento de Química da Universidade de Aveiro

Professora Doutora Margarida Sâncio da Cruz Fardilha

Professora Auxiliar Convidada da Secção Autónoma de Ciências da Saúde da Universidade de Aveiro

Professor Doutor Francisco Manuel Lemos Amado

Professor Associado do Departamento de Quimica e Director da Escola Superior de Saúde da Universidade de Aveiro

Agradecimentos

Venho por este meio expressar toda a minha gratidão a todas as pessoas mais próximas que durante esta batalha estiveram sempre presentes na minha vida e que de alguma forma não estão mencionados aqui.
Ao meu orientador o Doutor Francisco por toda as paciência, motivação, confiança, disponibilidade e enriquecimento de cultura geral em todas as reuniões e pelo seu apoio científico que sem duvida contribuiu bastante para o avançar deste projecto.
Ao meu co-orientador o Doutor Hugo por toda a disponibilidade e apoio ao longo deste projecto integrando-me no seu grupo de investigação e proporcionando-me oportunidades de aprendizagem não só na área da dissertação mas também na área de oncologia em geral.
Ao Doutor Rui Vitorino que apesar de não ser meu orientador posso dizer que se portou como tal ajudando e enriquecendo os meus conhecimentos científicos em toda a parte experimental deste trabalho.
A estas pessoas o meu muito obrigado pois sem duvida que foram os pilares de construção deste trabalho.
Ao Doutor Alexandre Ferreira pelo elo de ligação entre a Universidade de Aveiro e o IPO do Porto.
A toda a equipa do IPO, da Virologia, da Oncologia Molecular: à Marlene, à Carla, à Margarida, à Sandra, à Joana Dias, à Andreia, à Isabel e em especial à Joana Ribeiro por todos as duvidas esclarecidas na área e por todos os momentos de boa disposição que tornaram este trabalho mais agradável e que mais do que colegas se tornaram amigas.
Ao Doutor Eurico Monteiro, ao Doutor Eduardo Breda e à enfermeira Anabela Machado do serviço e Otorrinolaringologia do IPO por toda a ajuda e disponibilidade nas consultas com os pacientes ao longo do projecto.
A toda a equipa da Liga Portuguesa Contra ao Cancro em especial ao Doutor Leite Moreira e ao Doutor Vidal por toda a ajuda na angariação de controlos para este estudo e por todos os momentos de sabedoria e boa disposição.
Um especial agradecimento aos meus amigos mais presentes: à Jô, à Andreia, à Catarina, à Su, à Silvana, à Rafa, à Claudia, , à Rute, à Silvia, à Denise, à Joana S., à Ana, à Ana L, ao David, ao Toy, ao Capitão, ao Luis, e ao Jason. Ao Simmy e ao Paulo que para além da amizade ainda deram umas ajudas no excel. Obrigado por todos os bons e maus momentos e por me ajudarem a evoluir pessoalmente.
Por fim e sempre mais importante que tudo no mundo à minha família por me ajudarem a saber lutar sempre por mais e por estarem sempre presentes incondicionalmente para mim.
Um especial beijinho aos manos Eurico e Ricardo e à mana mais tardia Kate.

"A vida é maravilhosa quando não se tem medo dela."

Charles Chaplin

keywords

Epstein-Barr, Human papillomavirus, saliva, head and neck cancer, virus, tobacco, alcohol, proteomic, peptidomic, cell-cell interaction, aminoacids, proteolysis, fingerprint

abstract

After irreversible mutations in a cell, this can enter in a tumourogenese state evolving from a normal state to a hyperplasic originating a *carcinoma in situ* followed by an invasive carcinoma. There is also different aetiologic agents like tobacco, alcohol, a deficient diet and oral hygiene or viral infections that will promote the occurrence of mutations and the development of a possible invasive carcinoma. Head and neck cancers represent 5% of all cancers in Portugal with a mortality rate of about 50%, being a huge problem in public health. This way it becomes important to discover a new diagnosis method to discover and identify head and neck cancers as soon as possible to provide an effective treatment.

Saliva, composed by proteins have been recently investigated and proved to be a very useful sample in the detection of different types of diseases when associated with Molecular Biology techniques. Recent proteomic and peptidomic analysis showed that increased expression levels of proteins or proteases in saliva could indicate diferent health status.

This work addresses the first exploratory study regarding salivary proteases and peptidome with head and neck cancer patients. Results lead us to think in each way animal viruses are implicated in some human cancers and if they’re latency state help in the development of cancer. Proteolytic analysis revealed that maybe it would be possible to detect larynx and oropharynx cancers in previous stages but we need further studies to comprise that.

Palavras-Chave

Epstein-Barr, Human papillomavirus, saliva, head and neck cancer, virus, tobacco, alcohol, fingerprint, proteomica, peptidomica, interação celular, aminoacidoss, proteolise

Resumo

Após mutações irreversíveis numa célula, esta pode entrar em num estado tumourogenese evoluindo de um estado normal para um hiperplasico originando um carcinoma in situ, seguido de um carcinoma invasivo. Existem também diferentes agentes etiológicos como o tabaco, álcool, uma dieta deficiente e higiene oral ou infecções virais que podem promover a ocorrência de mutações e o desenvolvimento de um possível carcinoma invasivo. O Cancro da cabeça e pescoço representa 5% de todos os cancros em Portugal, com uma taxa de mortalidade de cerca de 50%, sendo por isso um grande problema na saúde pública. Desta forma, torna-se importante descobrir um novo método de diagnóstico para detectar e identificar cancros da cabeça e do pescoço, o mais rapidamente possível para proporcionar um tratamento eficaz.

A saliva, composta por proteínas tem sido recentemente investigada e provou ser uma amostra muito útil na detecção de diferentes tipos de doenças, quando associado com técnicas de biologia molecular. Análises proteomicas e peptidomicas recentes mostraram que o aumento dos níveis de expressão de proteínas ou proteases na saliva pode indicar o estado de saúde em que se encontra o individuo.

Este trabalho aborda o primeiro estudo exploratório sobre Proteinas e péptidos salivares com pacientes com cancro de cabeça e pescoço. Os resultados nos levam-nos a indagar se os vírus animais estão implicados de alguma forma neste tipo de cancros e se eles apesar estado de latência ajudam no desenvolvimento do cancro. A análise protolitica revelou que talvez fosse possível detectar cancros da laringe e orofaringe em estádios iniciais, no entanto precisamos de mais estudos que verifiquem estes dados.

Index

Abbreviators list

2-DGE – 2 Dimension Gel Electrophoresis

aPRP – acidic Proline-rich Proteins

bp – base pair

bPRP – basics Proline-rich Proteins

E – Early region

EBNA – Epstein-Barr Nuclear Antigen

EBV – Epstein-Barr virus

FISH - Fluorescent In Situ Hibridization

gPRP – glycosylated Proline-rich Proteins

HNC – Head and Neck Cancer

HNSCC – Head and Neck Squamous Cell Carcinoma

HPV – Human pappiloma Virus

IARC – International Agency for Research in Cancer

IM – Infectious Mononucleosis

IPO Porto FG EPE – Portuguese Institute of Oncology of Oporto

L – Late region

LCR – Long Controlled Region

LMP – Latent Membrane Protein

ORF – Open Reading Frame

pRb – Retinoblastoma Protein

PRP - Proline-rich Proteins

SNPs – Single Nucleotide Polymorphisms

OSCC- Oral Squamous Cell Carcinoma

Index of figures

Index of Tables

Index of Graphics

1. CANCER PHYSIOPATHOLOGY

A normal cell in the body will grow and divide in a controlled way in order to produce more cells keeping the body healthy, in this manner when a cell became old or damaged this cell will initiate a programed dead and will be replaced with a new one, however, sometimes cells suffer small mutations in its genetic material that will affect the growing and division processes and that could lead to an uncontrolled cell proliferation (2).

Cancer is a multistep process characterized by an uncontrolled cell proliferation usually implicated by several changes in the genome, which usually lead to a recessive function of tumour suppressor genes or to a higher production of oncogenes that will transform normal cells into a progressive evolution from premalignant cells into invasive cancers(3)

Different types of damaged cells will induce different types of cancers but all types of cancer will have some common processes(2). In all cases tumorigenesis occurs when the anticancer mechanisms like apoptosis, programed cell death or DNA reparation fails. Scientific advances made over the past decades, both in animals and *in vitro* studies in different organs, suggest that tumorigenesis has common intermediate steps(3).

In 2000, Hanahan and Weinber suggested that tumorigenesis is controlled by six essential alterations that are common to all types of cancers which are: self-suffiency in growth signals, insensitivity to anti-growth signals, evading apoptosis, limitless replication, sustained angiogenesis and tissue invasion and metastasis – Figure 1(3).

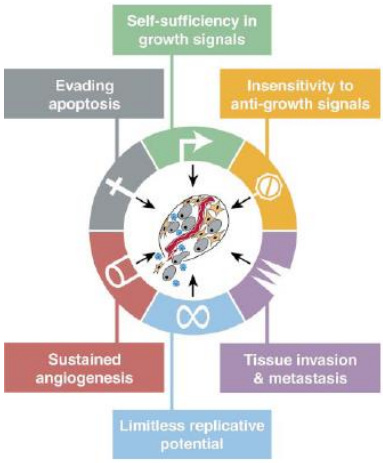


Figure 1 - Hallmarks of cancer - The six alterations proposed to the development of tumourigenesis. (Figure adapted from (3))

Usually these alterations are controlled by the maintenance genome under the control of complex mechanisms. Additionally, cell cycle checkpoints in mitosis assure the identity of the DNA information. Thus, investigators believed that tumour cells developed increased mutability due to genomic instability by decreasing the functions of the caretaker systems, like the p53 tumour suppressor protein (3). In this way when all of these systems tend to fail normal cells will grow in an uncontrolled way leading to possible cancer(2).

Later in 2011, progresses in cancer research added, to the previous six alterations in the cells, the ability to evade immune destruction and the capability to reprogram the cellular metabolism. In fact, it was also discovered that a tumour is not only a mass of proliferative cells, but a complex microenvironment composed by different cell types with heterotypic interactions - Figure 2(4).

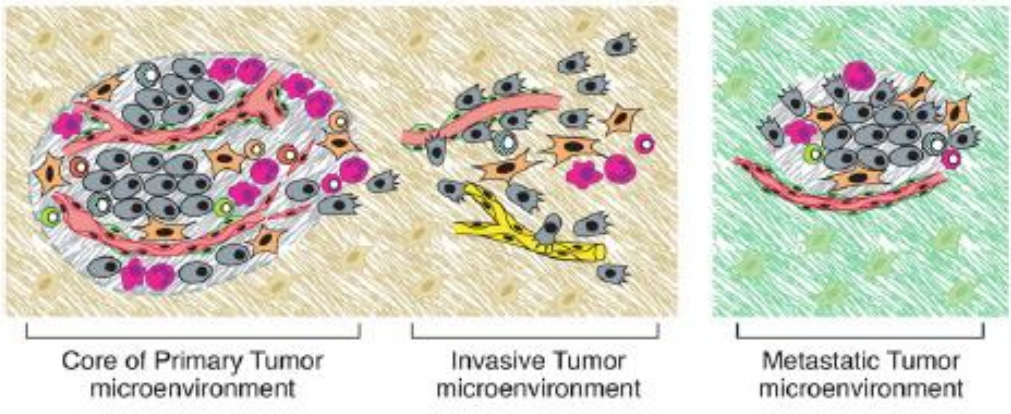


Figure 2 – Tumour Microenvironment - Example of a tumour microenvironment composed with different types of cells and different functions (figure from(4))

Different types of cancer will provoke different signs and symptoms in patients according with the site of the cancer, however most of the patients will experience common general signs and symptoms as fever, fatigue, weight loss, pain, skin changes, unusual bleeding and presence of new tissue masses(5).

Nowadays there are several tests like CT and MRI scans, biopsies of the tumour sites and blood tests for tumour markers specifics that allow the physicians to have a better diagnose of the stage, location and size of the tumours(6).

2. HEAD AND NECK CANCERS

Head and neck cancers include a diverse group of salivary gland neoplasms and squamous cell cancers from the mucosa's of the aero digestive tract, however there are some unique cancers in this region like nasopharyngeal carcinomas, Sino nasal undifferentiated carcinomas, esthesioneuroblastoma and thyroid cancers(5).

The majority of head and neck cancers - 95% - are squamous cell carcinomas (HNSCC) that occur in the mucosa and can be divided according six anatomic regions: oropharynx, hypopharynx, larynx, oral cavity, nasopharynx and paranasal sinus – Figure 3 and Table I (7).

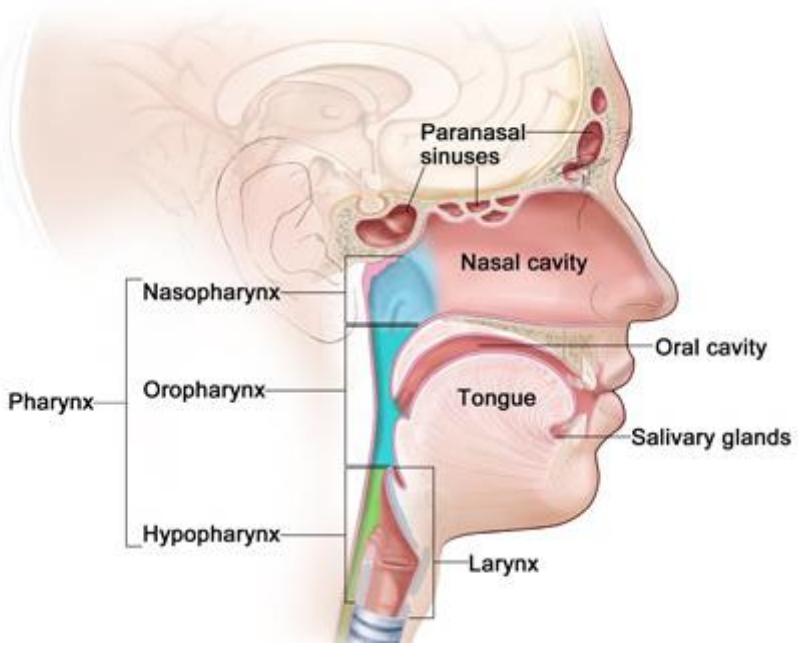


Figure 3 - Head and Neck anatomy

Specific signs and symptoms of this types of cancers includes hoarseness or change in voice, white or red patches in oral cavity, unusual bleeding or pain in the mouth, masses in head or neck area, persistent pain chewing or swallowing and persistent nasal obstruction or congestion(5).

Despite the low incidence rate, the high mortality rate is in most of the cases associated with the late diagnosis (8). In fact, most of the diagnosed cases are invasive tumours involving cervical lymph nodes and in some cases with local or distant metastases (9).

Table I- Different types of cancers from head and neck (only nasopharynx cancer is not Squamous Cell Carcinoma) (Adapted from ())

Different types of head and neck cancers	
Oral Cavity	Floor of the mouth
	Tongue
	Lips
	Buccal surfaces
	Gums
	Hard Palate
Oropharynx	Papillae
	Tonsils
	Epiglottis
	Pharyngeal Walls
	Soft Palate
Hypopharynx	Piriform sinus
	Lateral and Posterior pharyngeal walls
	Posterior faces of larynx
Larynx	Vocal Chords
	Larynx
	Subglottic and Supraglottic
	Thyroid
	Cricoid and arytenoid cartilages
Paranasal sinuses	Maxilla
	Ethmoid, sphenoid and frontal sinuses
	Nasal Cavity
Nasopharynx	Mucosal surfaces and structures of the cavity behind nasal passages

2.1 Epidemiology

According to the International Agency for Research on Cancer (IARC), head and neck cancers (HNC) account for about 5% of all of the human cancers in Portugal has we can see in Table II. These data show that despite the relative low incidence, there is a higher incidence in men when compared to women. Some results are observed for mortality rates, despite the low incidence almost 50% of the patients die and therefore this is considered an important public health problem(8).

Table II - Incidence and Mortality rates by gender in Portugal for Head and Neck cancers according to IARC (4).

Cancer	Incidence (number of cases per 100,000)			Mortality (number of cases per 100,000)		
	Men	***Women***	***Both gender***	***Men***	***Women***	***Both gender***
Lip, oral cavity	3,2	1,3	2,4	1,9	0,7	1,4
Nasopharynx	0,3	0,2	0,3	0,3	0,1	0,2
Other pharynx	1,7	0,1	1,0	2,1	0,2	1,3
Larynx	2,4	0,1	1,4	2,7	0,2	1,7

2.2 Aetiology

IARC identified almost 500 agents that are associated with the development of cancer in humans, which can be divided in two different groups according to their hazard risk. These agents were selected based on evidence of human exposure and evidence or suspicion of carcinogenity(10).

The recent evolution of genetics has lead to the development of tools that could help clinicians to detect, diagnose and predict the development and the behaviour of a specific cancer. Recent molecular techniques have provide us a better knowledge on the carcinogenesis on HNSCC where it's described several epigenetic and genetic alterations in invasive cancers – Figure 4 (11).

Head and neck cancer is considered to be multifactorial and its development is associated with several chromosomal abnormalities besides other co-factors related with behavioural risks that will induce damages in DNA such as: alcohol and tobacco(10), chewing of betel quid (seeds of the betel palm and tobacco wrapped in betel leaf), diet deficiencies (12) oral hygiene and viral infections (13), salted fish and smoking wood dust(10).

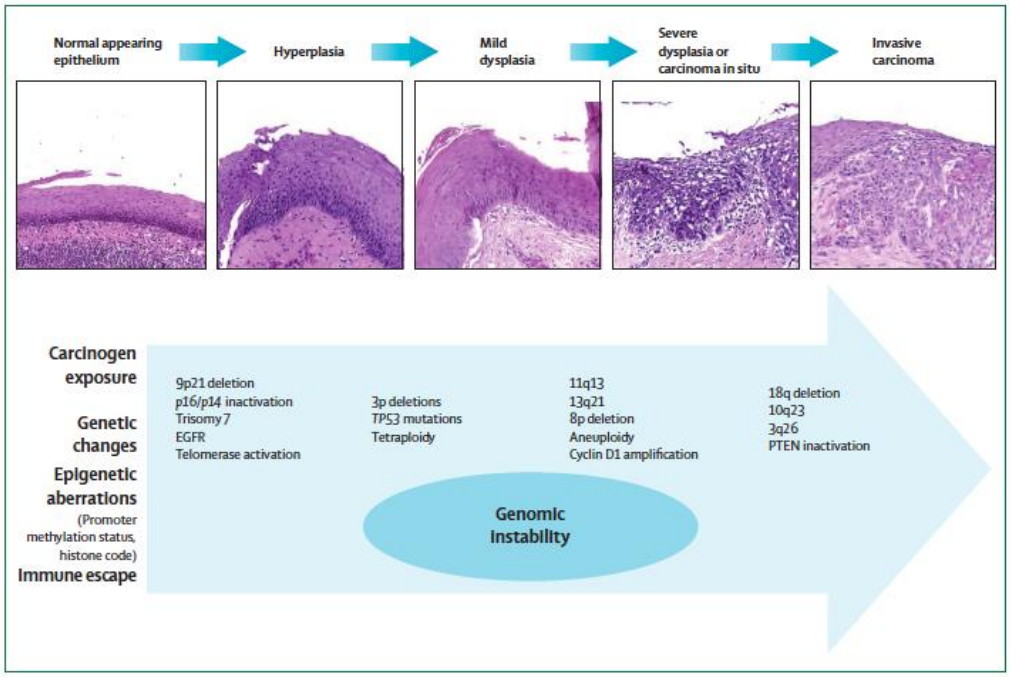


Figure 4 - progression and molecular changes in head and neck cancers that suspects to transform a normal epithelium into an invasive carcinoma (figure from (10)

3. RISK FACTORS

Nowadays there are several external factors to the human being that can increase the probability of an individual to develop cancer, despite this factors influence in the cancer development they do not cause cancer directly once there is several persons exposed to several risk factors that never develop any type of tumour or neoplasia. **REF head and neck: squamous cell carcinoma: an overview**

One of the best ways to prevent head and neck cancer is to avoid the recognised risk factors and to visit a dentist regularly since examination in the oral cavity can identify pre-cancerous lesions that will allow the physicians an easier treatment than a stage 3 or stage 4 cancers(12, 14).

3.1 Alcohol and Tobacco consumption as Risk factors

Tobacco and alcohol are the major environmental risk factors for head and neck cancers due to their mutagenic effect(15). Tobacco is composed by nitrosamines (N-nitrosonornicotine (NNN), polycyclic aromatic hydrocarbons (PAH) and 4-

(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK)) (12) type I that are considered to be carcinogenic agents(10). Moreover alcohol is metabolized in acetaldehyde by alcohol dehydrogenase and acetaldehyde has proved to have a mutagenic effect on DNA(16). However literature has shown that tobacco and alcohol carcinogenic effect is dose-dependent, and consumption in small doses might not increase the risk of oral cancer substantially(12).

The consumption of alcohol and the exposure to tobacco in high doses promotes damages in the DNA that can induce the development of cancer. Recent studies with spectral karyotyping, comparative genomic hybridization and fluorescent in situ hybridization (17) have shown several chromosomal alterations in HNSCC patients (gain of 1q, 3q, 5p, 7q, 8q, 9q, 11q, 12p, 14q e 15q; and loss of 1p, 3p, 4p, 5q, 8p, 10p, 11q, 15q and 18q) (18-21).

3.2 Viral mechanisms as Risk Factors

Virus can be defined as an intracellular parasite with extracellular phase dependent on the host cell functions and metabolism and with DNA or RNA genome encased in a protective shell.

Generally viruses cycle is composed by three stages: the infection, in which the virus penetrates the cell trough the attachment mediation of the viral surface proteins to the cellular receptors; the replication, in which a viral protein inactivates proteins like p53 or Rb in preventing cell apoptosis and the lysis in which there is a cell rupture and the release of new replicated viruses (22).

Virus and parasites contributes to oncogenesis accelerating mutations rates and enhancing proliferative signals through inflammation. Several viruses with different transmissions are associated with different types of cancers – Table III(23).

Table III - Correlation between different viruses and different types of cancers. The cancers in parentheses are a possible correlations. (Figure from(23).

Virus	Cancer	Transmission
HPV	Cervical, penile, rectal, oropharyngeal, Head and Neck	Sex, transiently by saliva
HBV	Hepatocellular [cholangiocarcinoma]	Sex, breast milk
HCV	Hepatocellular [cholangiocarcinoma]	Sex
EBV	Burkitt's lymphoma, nasopharngeal, posttransplant proliferative disease [acute lymphoblastic leukaemia, breast, gastric carcinoma, ovarian]	Sex, saliva
Cytomegalovirus	[Brain]	Saliva, sex
HTLV-1	Adult T cell leukaemia	Sex, breast milk
Merkel cell polyomavirus	Merkel cell cancer	Probably saliva

Off all viral infections, the most frequent carcinogenic viruses are Epstein-Barr virus (EBV), and human papilloma virus (HPV)(13). According to recent studies EBV contributes to oncogenesis and is commonly associated with B-cells diseases and usually detected in nasopharyngeal carcinomas, Hodgkin's disease, Burkitt's lymphoma and post-transplant B cell lymphomas (24).

HPV is considered to be a potential risk factor for head and neck cancer. This virus can be observed in early tumorigenesis, and more recently it has been suggested that HPV is a marker of a favourable prognosis (25).

Human papilloma virus and head and neck cancer

In 1949, papillomaviruses were first reported in a human mole visualized by electron microscopy and later in 1963 its genome was unravelled. Studies have contributed to the theory presented in early 70's that papillomaviruses could be involved in the aetiology of cervical cancer (26). Later in 1980s it was possible to isolate HPV types 6 and 11 from genital warts and HPV types 16 and 18 from cervical cancer biopsies confirming the role of HPV in the development of these lesions. Recent data demonstrate that HPV 16 is linked to other widespread human tumour types as the oropharyngeal cancer (26).

Structure and regulation of HPV

The HPV is a double stranded DNA virus and his genome contains 7200 to 8000 base pairs (17). The viral particle is approximately 55 nm and is composed by a capsid composed by 72 capsomeres (histone-like proteins): the major capsid protein is encoded by L1, a protein highly conserved amongst the different types of papillomavirus; and the L2 that is more variable(26).

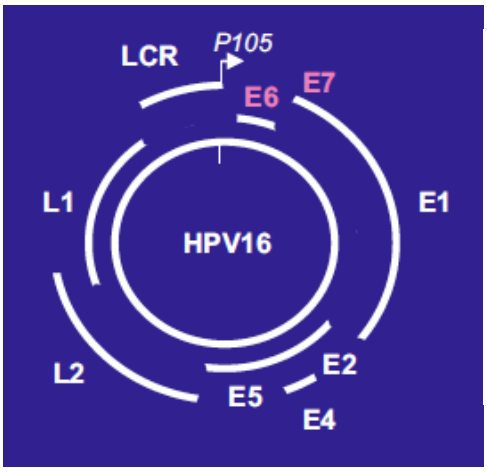


Figure 5 - HPV Genome HPV genome - It is possible to distinguish the different viral proteins in Early and Late regions and the Long Controlled Region (figure from(1))

Most of the *papilloma virus* sequences can be found in online databases such as EMBL and GenBank(27).

The genome of HPV can be divided in three different regions; an early region (E) from E1 to E7 responsible for the lytic cycle, DNA replication and regulatory

functions; a late region (L) divided in L1 and L2 that encodes structural proteins and compose the viral capsid and a long controlled region (LCR) also known as upstream regulatory region positioned between E6 and L1 as showed in figure 5 (26, 27).

According to the oncogenic activity of HPVs these viruses were classified into different groups(28): low-risk HPVs (type 6, 11, 13, 40, 42, 43, 44, 54, 61, 70, 72, 81 and 89); in probable high-risk HPVs (type 26, 53, 66, 68, 73 and 82); in high-risk HPVs (type 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59); and in undetermined-risk HPVs (type 30, 32, 34, 62, 67, 69, 71, 74, 83, 84, 85, 86, 87, 90 and 91)(29).

Human Papillomavirus viral proteins

The early region of HPV genome is responsible for the regulator and replication functions of HPV proliferative cycle. This region encodes for two regulatory proteins, E1 and E2 that control replication and transcription and three oncogenes, E5, E6 and E7 that control the transformation process(26, 27).

E1 represents the most conservative structure between different types of papillomavirus since it is essential for viral replication(27). E1 has binding functions for site-specific DNA sequences and shows helicase activity dependent of ATP required in the viral DNA replication(26).

E2 encodes for a protein that will provide an intragenomic regulation of viral gene expression by generating dimers at specific binding sites. E2 interacts with E1 by stimulating the replication enhancing the binding between the origin of replication and E1(26). Deletions in this ORF are highly correlated with the transition of human cells into a malignant state. In fact, premalignant lesions do not appear to have E2 expressed and moreover a mutation in the E2 ORF within the viral LCR has shown to contribute for cellular immortalization(26).

The E4 protein is originated from a single splice sequence between E1 ORF and E4 ORF that originates the main transcripts in HPV induced lesions(26). *In vitro* studies have shown that E4 interact with the keratin cytoskeleton inducing the collapse of this network contributing for cellular indifferentiation(30, 31).

E6 and E7 are capable of coding growth-stimulating proteins, this way these genes are often found in malignant tumours and are also capable of tissue immortalization in cultured cells (26). These proteins are considered the transforming proteins of HPV and are responsible for the inactivation of some of the most important cell cycle regulators, thus leading to cell proliferation. These events are often observed in the majority of HPV-associated cancers and in vitro studies have shown that this mechanism is the essential route for HPV-associated carcinogenesis(15, 26, 32).

E6 is composed by 151 amino acids and two zinc finger structures formed by four Cys-X-X-Cys motifs, capable of binding to p53 and promoting its degradation by the ubiquitin proteasome pathway(26, 33). This interaction between E6 and p53 protein causes a genetic instability that prevents cells to enter apoptosis and activate repair mechanisms(1, 26). Moreover E7 is able to bind the Retinoblastoma protein (pRb) and therefore releases the transcription factor E2F that is required for cell proliferation. Hence, the G1 arrest is avoided and cells start to activate the proliferation pathway(15, 26, 32).

Epstein Barr virus and head and neck cancer

In 1958, Denis Burkitt theorized that the etiological agent responsible by a cancer that was affecting children in several regions of Africa could be a virus(24), later this cancer was designated as the Burkitt's lymphoma (BL) (34) and in 1964 Epstein, Achong and Barr detected by electron microscopy herpes virus-like particles in a cell line obtained from a Burkitt's lymphoma biopsy and name it as Epstein-Barr virus(24). Afterwards, studies revealed that the sera from patients with BL had higher antibody titters to EBV than controls, since then, scientific community has developed a large number of studies to clarify the role of EBV in the aetiology of different diseases(24, 34). The carcinogenic potential of EBV was showed in vitro by its ability to transform resting B-cells(24) and by inducing lymphomas in cotton-top marmosets and in owl monkeys(34).

In fact it is known that EBV infection is the cause of infectious mononucleosis (IM) and is also associated with several malignancies, such as undifferentiated nasopharyngeal carcinoma (UNPC), T-cells lymphoma and immunoblastic lymphomas(24, 35).

Structure and regulation of Epstein Barr Virus

Epstein Barr virus (EBV) is a *gammaherpesvirus* from the Lymphocryptovirus genus(24) composed by a linear double stranded DNA with 184 kbp(34). EBV is composed by a glycoprotein spiked outer envelope, a protein between the envelope and the nucleocapsid, and a nucleocapsid that involves the viral genome in a toroid-shaped protein core like the other herpesviruses(34).

After infection, the viral DNA converts into a circular episome with a specific number of terminal repeats conserved from the parental genome. Its viral genome presents a series of 0,5kbp of terminal direct repeats and internal repeats sequences, which divide the genome into short and long domains that affect genetic transcription(34, 36).

EBV has the ability to produce over 85 proteins, however only some of them are well known and have a clear role in infection and transcription(34). From all the

EBV proteins the most important are the EBNAs (EBNA-1, EBNA-2, EBNA-3a, EBNA-3b, EBNA-3C, EBNA-LP) and the LMPs (LMP-1, LMP2-A and LMP-2B) that are important for cellular control and transcription(34, 36). There are mainly two EBV subtypes that usually infect humans, the EBV-1 and the EBV-2, which differ in the genetic organization of the nuclear antigen (EBNA-2, EBNA-3a, EBNA-3b and EBNA-3c)(34, 35).

Infection by EBV occur mainly in childhood(24), usually transmitted by saliva from host to host(34). This infection is not fully understood but primarily begins when the oropharyngeal epithelium cells are infected and occur virus replication, due to the proximity with the oropharyngeal epithelium the B-lymphocytes will be infected and will present a latent state(34).

An in vitro model of EBV infection and transformation in lymphoblastic cell lines showed high expression of some B-cell activation markers revealing that these cells were activated by antigenic or mitogenic simulation(24). In fact studies suggest that EBV-induced immortalization can be caused by activation of cellular pathways of the B-cells proliferation(24, 34, 36).

Viral proteins in Epstein Barr virus

EBV encodes for a series of latent genes essentials for cells immortalization and with different functions like replication, maintenance, transcription or signalization(34).

EBNA-1 has a fundamental role in the maintenance of latent infection, this gene is responsible for replication of EBV genome and segregation of viral episomes during the mitosis and is constituted by DNA sequence specified in binding the phosphoprotein in the origin plasmid replication(34).

EBNA-2 is responsible for the distinction between EBV-1 and EBV-2 types. This transcription factor is necessary for cell immortalization and is one of the first to be detected after EBV infection(34).

EBNA-3 is encoded by three genes that connect to the viral genome and form the different types of EBNA-3, the EBNA-3A, the EBNA-3B and the EBNA-3C, all of

the three genes are transcriptional regulators but only EBNA-3A and EBNA-3C are essential for B-cells transformation and EBV immortalization. Despite the function of EBNA-3B is not well known its known that this gene interacts with CBF1 like EBNA-3C. EBNA-3C also increases the production of LMP1 and overtakes the pRb checkpoints in the cell cycle(34).

EBNA-LP is also one of the first viral proteins to be produced after EBV infection. This gene is composed mainly by RNA splicing sequences where its ORF is constituted by two exons, the Y1 and the Y2. EBNA-LP contributes to immortalization and also interacts with transcriptional factors and inactivates p53 and pRb(34).

LMP-1 act as constitutively active receptor (CD40) and mimics this signal by association with the tumour receptor associated factors. This protein can recruit any gene and inhibit apoptosis by increasing the levels of *Bcl-2* and *a-20*, this way it is responsible by the signalization and the immortalization of the cell(34).

LMP-2 encodes for LMP-2A And LMP-2B, these proteins are involved in the beginning of the latent state of EBV and may be associated with nasopharyngeal carcinomas and Hodgkin's disease but this relation with the oncogene of these diseases is yet unknown(34).

4. PROTEOMICS AND HEAD AND NECK CARNCERS

Since Marc Wilkins introduced the word proteome in 1995 great advances have been made in the fields of proteomics(37).

Proteomics can be defined by the identification and detection of proteins based on the study of structure, concentration, function, regulation and dynamic interactions of proteins (38), these complex information's can reveal different post-translational modifications (PTMs), proteolitic clivages and alternative splicing that can be later related with different types of diseases (39).

With this advances it become important to map all of the proteins, with that purpose it was born the Human Proteome Project (HPP) launched by the Human Proteome Organization (HUPO) (40, 41). This project was originated with the aim to map the proteins functional network in different times and conditions in a particular organism, tissue or cell providing this way new information and tools for a better understanding in health and disease control (41).

The integration of new research instruments more advanced and sensitive than 2D electrophoresis such has Mass Spectrometry (MS) techniques and bioinformatics allowed researchers to identify more than 4000 proteins in a particular shotgun run (42) and generate different databases from different organisms(43-45).

Peptides and proteins can be divided in different platforms in which they are characterized and classified according to qualitative and quantitative point of view (46, 47).

In a qualitative platform analysis the all set of peptides, proteins and PTMs present in a sample are defined despite the different proteins concentration distribution, this approach has disadvantages, for instance in the case of plasma only after the depletion of the most abundant proteins the low ones will be revealed (48).

The need to complete this platform with quantitative information was confirmed when researchers found out that the concentration of some proteins or some PTMs could be associated with specific diseases and these ones could be eligible as new biomarkers (46). Quantitatively this platform can be divided in relative and absolute quantification. In a relative quantification it is possible to compare the differences

between a healthy versus a pathological proteome subject by statistical approaches analysing increases or decreases in protein levels (46). Recent studies also demonstrate that increased or unique levels of specific proteins may be associated with pathological disorders and many viral diseases(49-51).

Proteomics technology have also been used as a screening tool due to its sensitivity and specificity, to diagnose and early detect cancer disease, cancer-specific alterations in protein abundance or in PTMs like glycosylation or phosphorylation that can be developed as early disease biomarkers(52, 53).

An ideal cancer biomarker can be a protein, RNA, cytokine, carbohydrate or chemokine and should have a high analytical sensitivity and specificity that allow an easy and reliable analysis (25, 54).

Several proteins can be differently expressed or secreted comparing healthy controls with patients with cancer providing different protein expression signatures that allow researchers to identify new biomarkers (55, 56). When a new protein biomarker is identified new bioinformatics techniques like protein chips with specific modified chromatographic surfaces and a specific protein pattern acquired by the MS technique can be developed as a diagnostic procedures when combined with a precise algorithm that allow to classify the sample as unclassifiable, healthy or disease affected patient (57).

5. SALIVARY PROTEOMICS

Numerous studies in genomic and proteomic fields in the last 10 years lead to the discover of several biomarkers that help the clinicians to predict and evaluate the prognosis of a cancer, this way clinicians can evaluate the predisposition of a patient to a specific cancer and screen it in a early phase in a manner to improve the clinical outcome(58).

Recent studies combining the new MS proteomic techniques with quantitative and qualitative proteomics approaches have provided the capability to identify large number of proteins from different biological fluids samples such as plasma/serum, urine, cerebrospinal fluid, bronchoalveolar lavage fluid, synovial fluid, nipple aspirate fluid, tear fluid, amniotic fluid and saliva (58, 59)

Saliva has proven to be an appealing sample in proteomic research and clinical applications due to its salivary proteome that could provide a diagnosis for several diseases trough MS analysis (60-62).

The profiling of these proteins with high sensitivity and specificity could indicate different hereditary diseases and autoimmune diseases like Sjögren's syndrome, systemic sclerosis and diabetes, several malignancies and cancers and several oral diseases like periodontitis and dental caries allowing a better diagnosis and treatment by the clinicians(55, 63-68).

Whereas approximately 20% of these salivary proteins can also be found in human plasma with similar functions and disease-linkage, recent studies demonstrate that increased or unique levels of specific proteins may be associated with pathological disorders such as immunodeficiency virus or systematic diseases(49, 50, 55, 69).

In fact, the high sensitivity and specificity of these methods could indicate different stages of several diseases or give information about body conditions by measuring systemic levels of drugs in case of illicit drug use or measure the immune response to viral or bacterial infections allowing the clinicians to provide a better diagnosis and treatment (69-71).

Saliva analysis can be divided in the analysis of the whole saliva for disease diagnosis or in the analysis of gland-specific saliva to analyse a specific pathology in a specific major gland(69).

5.1 Composition of Saliva

Saliva is a biofluid that not only serves to lubricate and protect the oral cavity or to help in the digestion but has also functions that will maintain the mouth homeostasis(72, 73). Besides that, this biofluid also offer distinctive advantages in diseases diagnosis when compared to other biologic fluids, the fact of this fluid doesn't require special collection equipment's or the fact of its low cost and non-invasiveness makes of this fluid a valuable asset in the diagnosis of diseases specially in patients like children or older adults(74).

Human Saliva is produced in the oral cavity by three major salivary glands; parotid, sublingual and submandibular and several minor glands. Specific collection from each individual gland could provide information about gland-specific pathologies like infections, however, after the production of saliva from submandibular glands and sublingual glands the saliva from both glands enters the oral cavity trough the Wharton's duct making the separate collection from this glands a difficult process(75) – Figure 6 .

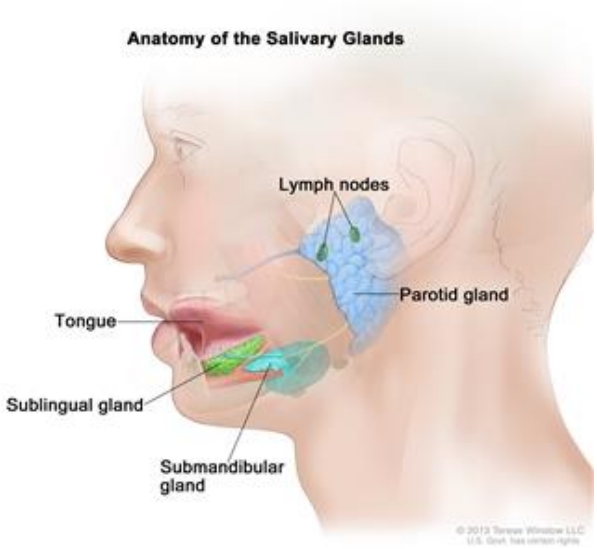


Figure 6 - Salivary glands anatomy

Whole saliva is a mixture from the secretions of the different glands and from non-salivary constituents such as expectorated bronchial and nasal secretions, serum and

blood derivatives, cellular components and epithelial cells, gingival crevicular fluid which is composed by bacterial and viral products and food debris (76-78).

Water, several electrolytes (sodium, calcium, potassium, magnesium, phosphates and bicarbonate), mucins, immunoglobulins, nitrogenous products, enzymes, nucleic acids, lipids, peptides and proteins mainly compose this biofluid(79, 80). All of these components have specific functions: bicarbonates and phosphates are important in the modulation of the pH in order to maintain the buffering capacity of saliva; mucins and macromolecule proteins contributes to the homeostasis maintenance of the dental plaque metabolism; phosphate calcium and proteins work in the modulation of demineralization and remineralisation processes and immunoglobulins, enzymes and proteins have antibacterial action (80-83).

Peptides and proteins present in saliva are multifunctional (81) and can be classified in five main groups; histatins, cystatins, proline-rich proteins (PRP), defensins and statherins – Table IV. (84) The proline-rich proteins can be subdivided in basics proline-rich proteins (bPRP), acidic proline-rich proteins (aPRP) and glycosylated proline rich-proteins (gPRP)(85).

Table IV - Salivary proteins - Origin and functions of most abundant salivary proteins (adapted from (107))

Protein Family	Functions	Origin
Alpha Amylases	Antibacterial, digestion, tissue coating	Parotid; Submaxillary; Sublingual
Acidic PRPs	Lubrication, mineralization, tissue coating	Parotid; Submaxillary; Sublingual
Basic PRPs	Binding of tannins, tissue coating	Parotid
Basic Glycosylated PRPs	Antiviral, lubrication	Parotid
Carbonic anhydrase VI	Buffering, taste	Parotid; Submaxillary
Cystatins	Antibacterial, antiviral, mineralization, tissue coating	Parotid; Submaxillary; Sublingual
Histatins	Antifungal, antibacterial, mineralization, wound-healing	Parotid; Submaxillary; Sublingual
Lactoferrin	Antibacterial, antifungal, antiviral, innate immune response	All salivary glands
Lysozyme	Antibacterial	Parotid; Submaxillary
Mucins	Antibacterial, antiviral, digestion, lubrication, tissue coating	All salivary glands
Peroxidases	Antibacterial	Parotid; Submaxillary
Statherins	Inhibition crystal form, lubrication, mineralization, tissue coating	Parotid; Submaxillary; Sublingual
Albumin	Carrier protein	Gingival crevicular fluid
Alpha Defensins	Antibacterial, antifungal	Gingival crevicular fluid
S100 A proteins	Different functions	Various

Beta Thymosins	Different functions	Gingival crevicular fluid

Since its composed by a large spectrum of proteins, recent proteomic technologies have been use to analyse and identify the salivary proteome quantitatively and qualitatively(69).

In the past years researchers have been catalogue the human proteins in the whole saliva using several MS platforms in order to understand the peptide/protein composition of this fluid(86). 2100 intact proteins from the aqueous supernatant of whole saliva were identified in one study allowing the researchers to discover that exist a significant overlap of proteins when compared to human serum proteome(86, 87). The insoluble part of the saliva also contains proteins; these proteins are mainly from exosomes and membrane vesicles secreted from cells presents in other body fluids(88).

With the development of MS databases it became important to map the structures and functions of the primary sequences of the proteins highly regulated by post-translation modifications (PTMs), some examples of this modifications are glycosylated or phosphorylated proteins and they allow to discover news hints in protein dysfunction when submitted to risk factors(89).

Besides the intact proteins and the PTMs whole saliva also have present low molecular weight proteins – peptides – with many biological functions. The salivary peptidome can also be used as diagnostic biomarker, is moderately stable to degradation and will complete the information collected from the intact proteins and the PTMs(89).

Several studies also indicate noticeable differences in the salivary proteome when comparing samples from children with samples from adults (90, 91) this way in our study we pretend to choose controls with the same year ranges than our patients.

5.2 Salivary proteome and Head and Neck Cancer

Clinicians are constantly searching for clinical applications that could monitoring a disease status, its the beginning, its progressions and its treatment outcome, for this propose they use a series of pre-requisites such as a specific biomarker for the disease in study, a technique for the biomarker study that does not involve an invasive approach and the use of technologies that allow to distinguish the different biomarkers(69).

As seen before saliva is a body fluid present in the oral cavity mainly composed by water (99,5%), proteins (0,3%) and inorganic substances (0,2%)(92) and its relationship with plasma levels makes of this easy to obtained body fluid with non-invasive nature and low cost processing an attractive diagnostic tool (86, 92). Despite the amount and composition of this body fluid depend patience health status(84) and according to many factors such as age, gender, blood type, size of the salivary glands; scientists are searching biomarkers in human saliva proteome for early detection of oral cancers (86, 93-95).

The primary reason for the high mortality rate in oral cancers is mainly the delayed detection once most of the patients present symptoms in a late stage provoking a higher recurrence rate after the first treatment(86). This fact supports the need for biomarkers with high sensitivity, specificity, and functionality in the detection of early stage head and neck cancers(69, 86).

Detection of differentially expressed proteins trough a mass spectrometry method revealed cytokeratin 4 and 13, zinc finger proteins, P53 pathway proteins, annexin A1 and beta and gamma actins as potential biomarkers for HNSCC and revealed also high levels of salivary transferrin in oral squamous cell carcinomas (OSCC) patients which proved to be highly specific and sensitive in the detection of OSCC(69). Another MS study validated by immunoassays also found that the combination of increased levels of S100A9, profiling, CD59, catalase and MBP2 could detect OSCC with 90% of sensitivity and 83% of specificity (86)

Different techniques as Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), Quantitative real time (RT-PCR), High performance liquid chromatography (HPLC) and DNA and RNA microarrays provide also potential biomarkers in the early detection of OSCC such as; albumin, endothelin-1,

Immunoglobulin heavy chain constant region gamma, transthyretin and also transferrin(96).

6. Viral Proteomics in Head and Neck cancers

Saliva contains a vast number of proteins that researchers believe to be involved in extensive variety of biological functions, metabolites as well as microRNA and mRNA transcripts (97-99).

The increasing of sensitivity in the genomic and metabolomics improved different methods of virus diagnostic in saliva samples allowing the association of EBV to nasopharyngeal carcinoma which according to literature presents higher levels of EBV DNA in stages 3 and 4 than in stages 1 and 2 (100).

Notwithstanding peptides and proteins are vulnerable to degradation since they generally have short half-lives, protein stabilization methods allowed MS (101, 102) analysis and recent studies demonstrate that increased or unique concentration levels of specific proteins may be associated with pathological disorders, mainly viral diseases(49-51).

Moreover this sample may be even more important to evaluate and following patients with viral infections of the oral cavity. In fact, many studies have been designed to identify tumour-associated markers in saliva(69, 103, 104). Despite the viral detection in samples not always allow a good correlation with clinic once a huge number of viruses is in the latent state in several types of tissues. This way the detection of viral markers can be a better approach for the identification of risk units(51). Since different types of peptides and proteins compose saliva we believe that is possible to identify viral peptides/proteins in this type of sample.

7. PROTEOMIC AND PEPTIDOMIC FINGERPRINT AS A POSSIBLE DIAGNOSE TOOL FOR HEAD AND NECK CANCERS

When saliva enters the oral cavity, protein salivary secretion will be vulnerable to a variety of biochemical and physiological processes promoted by a series of enzymes from the individual and from bacteria. This process will induce several post-translational modifications (105).

After the transcription and translation of the genes involved in the protein biosynthesis in the glands, several post-translational events like protein phosphorylation, glycosylation and proteolysis (exo and endo) will take place and the salivary proteins that enters in the oral environment will be formed with similar structural differences, originating different protein families (105, 106). Subsequently saliva enters the oral cavity and continues to suffer protein modifications that will lead to protein-protein complex formation, partial deglycosylation and a widespread of exogenous proteolitic cleavages until swallowing (105).

As we have seen before the major salivary proteins are secreted from different glands and will suffer different types of post-translational modifications. Several studies included in Castagnola review allowed the scientific community to be better elucidated about novel functions for different families of specific peptides and proteins(107). In table V is showed the most relevant post-translational modifications of salivary proteins.

Table V - Common Post translational modifications in Saliva proteins (adapted from 107)

Precursor Protein	Matures proteins	Post-translational modifications
Salivary acidic proline-rich phosphoprotein 1	Db-s, Pa, PIF-s, Pa 2-mer, Db-f, PIF-f, P-C peptide	Disulphide, further proteolytic cleavages, N-terminal pyro-Glu, phosphorylation, transglutamination
Salivary acidic proline-rich phosphoprotein 2	PRP-1, PRP-2, PRP-3, PRP-4, P-C peptide	Further proteolytic cleavages, N-terminal pyro-Glu, phosphorylation, transglutamination
Basic salivary proline-rich protein 1	II-2, P-E, IB-6	Further proteolytic cleavages, N-terminal pyro-Glu, N- and O-glycosylation, phosphorylation, transglutamination
Basic salivary proline-rich protein 2	IB-1, IB-7, IB-8a (Con1_/+), P-F, P-J, P-H	Further proteolytic cleavages, N-glycosylation, N-terminal pyro-Glu, phosphorylation, transglutamination
Basic salivary proline-rich protein 3	Proline-rich protein Gl 1-8	Disulphide (Gl 8), N- and O-glycosylation, N-terminal pyro-Glu, phosphorylation
Basic salivary proline-rich protein 4	Salivary proline-rich protein Po, protein N1, P-D,	N-glycosylation

	II-1, CD-IIg	
Hst 1 nonP	Hst 1, Hst 2	Further proteolytic cleavages, phosphorylation, sulfation
Hst 3	Hst 3, Hst 5, Hst 6	Further proteolytic cleavages
Mucin-5B precursor	Mucin-5B	N- and O-glycosylation, disulfide, phosphorylation
Mucin-19 precursor	Mucin-19	Disulphide
Mucin-7 precursor	Mucin-7 (MG2)	N- and O-glycosylation
Cyst SN precursor	Cyst SN	Disulphide, sulfoxide
Cyst SA precursor	Cyst SA	Disulphide
Cyst C precursor	Cyst C	O-glycosylation
Cyst S precursor	Cyst S	Disulphide, phosphorylation
Cyst D precursor	Cyst D	Disulphide, truncated forms
Statherin precursor	Statherin, statherin SV2	Phosphorylation, proteolytic cleavages, transglutamination
Proline-rich peptide P-B Precursor	Proline-rich peptide P-B	N-terminal pyro-Glu, proteolytic cleavages
a-amylase 1 precursor	a-amylase 1	Disulphide, N-glycosylation, N-terminal pyro-Glu, phosphorylation, proteolytic cleavages
Neutrophil Def 1 precursor	a-Def 1 and 2	Disulphide, phosphorylation
Neutrophil Def 3 precursor	a-Def 3 and 2	Disulphide, phosphorylation
Neutrophil Def 4 precursor	a-Def 4	Disulphide
Tb4 precursor	Tb4	Acetylation, phosphorylation,

		proteolytic cleavage, sulfoxide, transglutamination
Tb10 precursor	Tb10	Acetylation, phosphorylation, transglutamination
Protein S100A7 precursor	Protein S100A7	Acetylation, disulfide bond
Protein S100A9 precursor	Protein S100A9, protein S100A9 (short type)	Phosphorylation, sulfoxide

Since all of these processes suffers from the singleness of the individual, this information can be seen as fingerprint in which it is possible to link specific post-translational modifications to a specific pathology and determine in which way it affects or promotes its development (107).

Several MS studies with different types of stimulation and validated by molecular biology techniques, identified different biomarkers according to different types of head and neck cancers (107).

Pre cancerous and cancerous lesions of oral cavity revealed decreased levels of Statherins; oral cancers revealed increased levels of actin, myosin and albumin; HNSCC revealed increased levels of a-1-B-glycoprotein, complement factor B, a-1-acid-glycoprotein-2, fibronectin1, kininogen, Ig heavy chain C region gamma 2, transferrin, S100A9, fibrin b and cofilin 1; OSCC revealed increased levels of alpha-Def-1, M2BP, S100A9, protectin, catalase, profiling, Cyst SN and Transferrin (107).

8. PROTEOLITIC ACTIVITY IN HEAD AND NECK CANCERS

As we have seen before one of the constituents of human whole saliva are enzymes, some of these were already identified and they include catalase, amylase, carbonic anhydrase and proteases(108, 109). Alterations made by proteases results in smaller peptides and low-molecular-weight proteins (a significant fraction on the salivary proteome) unique for each individual(110).

Subsequently, the identification and characterization of the salivary proteases that modify structures and functions of salivary proteins involved in oral diseases (109) are a fundamental step to understand the oral microenvironment in proteolitic associated diseases or in health individuals(110). Furthermore it can also act as tool in clinical diagnosis and prognosis(110).

In spite of the fact that until today over 200 salivary peptides were identified, proteolitic cleavage events are not completely understood(109). However it is known that the majority of the peptides from the proteolytic fragmentation belong to PRPs, histatins and statherins(109).

With the view to understand the peptide formation trough proteolitic activity in saliva, several studies have been adopting different approaches and been relating small peptides to Cathepsin D, to trypsin, chymotrypsin or elastase activity (109). Cysteine peptidases like Cathepsins B, C, H and L, with strong anti bacterial and anti viral properties are inhibit by cystatins isoforms (S, C, D, SA, SN)(84).

Most of the peptides fragments found in saliva derive from PRPs, statherins and histatins there are extremely proteolyzed in oral cavity by proteases from epithelial cells, leukocytes from the gingival exudate, from the dental plaque and from several microorganisms(84).

Proteolytic activities may modulate functional activities or affect a disease state; this way the rate of proteolysis in saliva and in dental plaque could indicate dental caries vulnerability and periodontal disease(84).

9. OBJECTIVES

The present work can be described in two main goals:

First we aim to identify viral peptides/proteins in saliva prevenient from head and neck cancer patients and correlate the different peptides/proteins obtained from the saliva's from the patients with the presence or absence of EBV and HPV.

Second we aim to identify a peptidomic fingerprint of the proteome and the degradome of saliva's from patients with HNSCC.

Both approaches have the main objective to correlate these factors with different types of head and neck cancers in order to create a new diagnostic tool.

10. METHODOLOGY

10.1 Type of study

This work was preformed in the University of Aveiro and at the Molecular Oncology Group of the Portuguese Institute of Oncology of Oporto (IPO Porto FG EPE) with the collaboration of Virology Service from IPO Porto using the technical resources existent in the laboratories of these institutes.

This study was developed as a prospective case-control study with consecutive patients attended at IPO Porto FG EPE during the period of February to May of 2013 selected according with the criteria previously defined:

Inclusion Criteria:

- Patients with diagnosis of squamous cell carcinomas in the oral cavity and nasopharynx
- Histological confirmation of diagnosis
- Patients have to accept to provide samples of tissue, blood and saliva

Exclusion Criteria:

- Prior head and neck cancer
- Prior treatments of radiotherapy and/or chemotherapy
- Presence of blood in the saliva sample
- Patients that does not accept to participate in this study

Control Group

The control group was constituted by healthy individuals from Blood Donor Database at IPO Porto FG EPE with matched age and gender without history of cancer.

Ethical questions

This study did not interfered with the routine procedures decided by clinicians. Clinicopathological data was collected from individual clinical records and inserted on a database with unique codification. All procedures were submitted to approval

of the IPO Porto FG EPE Ethical Committee and all samples were obtained with the informed consent of the participants according to the Declaration of Helsinki.

11. MATERIAL AND METHODS

11.1 Saliva Samples

Saliva samples should provide us a wide range of comparison between the healthy subjects that should not exhibit any evidence of oral inflammations or pathologies (control group) and between the subjects of the disease in study (Study group)(17, 49).

These samples were collected according to standardized techniques in the Portuguese Institute of Oncology and will be kept refrigerated until their processing. Furthermore processed saliva was analysed in University of Aveiro with MS/MS techniques described further.

11.2 Patients and control groups

Samples from 82 patients with pathologies from oral cavity, pharynx and larynx areas were collected. From the samples collected 11 patients were excluded because they didn't fulfil the pathology inclusion criteria. 27 patients were excluded because they didn't have sufficient saliva sample for analysis.

Subjects included in the present study included 44 patients with head and neck carcinomas (confirmed histologically) with range ages between 34 and 72 divided in 3 groups: 24 in the Larynx group, which include Laryngeal and Vocal chords SCCs; 16 in the Pharynx + Oral Cavity group which include Floor of the mouth, Uvula and soft palate, tonsils, tongue, pharynx, pharyngolarynx, oropharynx and hypopharynx SCCs and 4 in the Nasopharynx + Nostrils group composed by nasopharynx and nostrils carcinomas.

The patients were followed up at Otorhinolaryngology Service at Portuguese Institute of Oncology Porto FG EPE and patients with prior head and neck cancers, prior treatments of surgery, radiotherapy and/or chemotherapy or samples with presence of blood in the saliva were excluded.

The HNSCC patients were compared with a control group composed by 10 healthy individuals without history of cancer and range ages between 47 and 72, collected in the North region of Portugal after a medical screening with internal physicians of Portuguese League Against Cancer. The clinical characteristics of enrolled subjects are summarized in Table VI.

Table VI - Clinical characterization of the study subjects

Patient	Sex	Age	Diagnosis	Group	Tobacco packages per year	Alcohol Grams per day
Control Group						
C1	M	59	-	-		
C2	F	64	-	-		
C3	M	47	-	-		
C4	M	60	-	-		
C5	M	64	-	-		
C6	M	58	-	-		
C7	M	57	-	-		
C8	F	72	-	-		
C9	M	59	-	-		
C10	M	51	-	-		
Study group						
1	M	67	Vocal Chords SCC	Larynx	13	36
2	F	72	Vocal Chords SCC	Larynx	No	No
3	M	44	Vocal Chords SCC	Larynx	53	26
4	M	59	Vocal Chords SCC	Larynx	40	52
5	M	55	Laryngeal SCC	Larynx	15	*

6	M	57	Laryngeal SCC	Larynx	74	*
7	M	67	Laryngeal SCC	Larynx	54	72
8	M	63	Laryngeal SCC	Larynx	No	144
9	M	47	Laryngeal SCC	Larynx	32	36
10	M	56	Laryngeal SCC	Larynx	60	108
11	M	60	Laryngeal SCC	Larynx	40	72
12	M	37	Laryngeal SCC	Larynx	39	72
13	M	55	Laryngeal SCC	Larynx	44	72
14	M	64	Laryngeal SCC	Larynx	No	72
15	M	59	Laryngeal SCC	Larynx	88	72
16	M	67	Laryngeal SCC	Larynx	-	-
17	M	46	Laryngeal SCC	Larynx	25	72
18	M	34	Laryngeal SCC	Larynx	15	102
19	M	53	Laryngeal SCC	Larynx	43,75	120
20	M	50	Laryngeal SCC	Larynx	30	72
21	M	55	Laryngeal SCC	Larynx	30	72
22	M	72	Laryngeal SCC	Larynx	200	72
23	M	48	Laryngeal SCC	Larynx	74	72
24	M	53	Laryngeal SCC	Larynx	32	26
25	F	87	Tongue SCC	Pharynx+ Oral Cavity	15	36
26	M	49	Oropharyngeal SCC	Pharynx+ Oral Cavity	30	No
27	M	67	Oropharyngeal SCC	Pharynx+ Oral Cavity	40	39
28	M	58	Oropharyngeal SCC	Pharynx+ Oral Cavity	96	72
29	M	57	Oropharyngeal SCC	Pharynx+ Oral Cavity	20	72
30	M	56	Mouth floor SCC	Pharynx+ Oral Cavity	61,5	*
31	M	39	Mouth floor SCC	Pharynx+ Oral Cavity	2,45	26
32	M	59	Uvula and Soft palate SCC	Pharynx+ Oral Cavity	44	72
33	M	52	Tonsil SCC	Pharynx+ Oral Cavity	45	181,75

34	M	49	Tonsil SCC	Pharynx+ Oral Cavity	35	300
35	M	54	Tonsil SCC	Pharynx+ Oral Cavity	40	48
36	M	50	Pharyngeal SCC	Pharynx+ Oral Cavity	45	117
37	M	51	Pharyngolaryngeal SCC	Pharynx+ Oral Cavity	39	*
38	M	63	Pharyngolaryngeal SCC	Pharynx+ Oral Cavity	80	72
39	M	48	Hypopharyngeal SCC	Pharynx+ Oral Cavity	-	-
40	M	57	Hypopharyngeal SCC	Pharynx+ Oral Cavity	40	**
41	M	46	Nostrils	Nasopharynx + Nostrils	No	**
42	M	52	Nostrils	Nasopharynx + Nostrils	30	36
43	M	63	Nostrils	Nasopharynx + Nostrils	No	**
44	M	34	Nasopharynx	Nasopharynx + Nostrils	No	**

(* former consumer; **sporadically consumer)

The present study was approved by the IPO Porto FG EPE Ethical Committee and all of the subjects included had been informed of the nature of the current research and had given their written informed consent according to Declaration of Helsinki.

11.3 Sample collection:

Saliva collection was performed to all of the study individuals after two hours of the patient last meal without parotid stimulation to a sterilized glass. This was followed by the addition of a 250µL 6M Urea solution per 1mL of saliva, divided 200µL factions and frozen at -80ºC until further analysis.

11.4 Sample preparation:

Fractionated saliva samples were submitted to centrifugation at 12000g for 10 min at 4ºC to remove possible food remains, cell debris and bacteria. Supernatant was collected adding 60 µL of 20% acetonitrile and resubmitted to centrifugation at 12000g for 10 min at 4ºC in case of visual precipitate after 5 min at room temperature. Spin filters (50 kDa cut-off) were conditioned with the supernatant and submitted to a 12000g centrifugation for 20 min at 4ºC and then 40µL of ammonia hydrogenocarbonate 50mM was added to the spin filters resubmitting it to a 12000g centrifugation for 20 min at 4ºC. Filtrates were then separated in one fraction of 100µL and two fractions of 50µL and the 50µL fractions posteriorly submitted to Speed Vac. All of the filtrates were stored at -80ºC for posterior peptide separation and identification.

11.5 Nano HPLC peptide separation:

Dried extracted peptides were resuspended in 5µL of a 0.3; 49.7; 50 trifluoracetic acid/water/acetonitrile (v:v:v) solution and submitted to ultrasounds 3 min. Samples were then spin centrifuged until 10000g and a 25µL solution of trifluoracetic acid 0.1% was added to the mix sample submitting it again to 3 min of ultrasounds and spin centrifugation.

Separation was preformed with 10µL of each sample in C18 column Pepmap (Dionex) in a Nano-HPLC 3000 Ultimate (Dionex/LC Packings, Sunnyvale, CA) and the fractions collected with Probot (Dionex/LC Packings, Sunnyvale, CA) directly by a laser desorption-ionization (MALDI).

MS/MS spectra were obtained in a 4800 MALDI-TOF/TOF Analyser (Applied Biosystems, Foster City, CA) and processed with Mascot software (v.2.1.0.4) for peptide/protein identification. Searches were made against Swissprot protein database with two different search approaches:

The first search approach was in *Virus* database with none enzyme selection, 40 ppm MS tolerance for precursor ions, 0.5 Da tolerance for fragment ions and with +1 peptide charge. Positive identification was considered for default significance threshold $p<0.05$ and individual ion scores above 44.

To estimate the false discovery rate (FDR), a database of decoy reverse database was created for all SwissProt resulting in 5 % FDR (false positives/ peptides (false positives + Total peptides)) × 100, only the peptides obtained from search FDR were considered for analysis.

Furthermore, all of the identified peptides were manually submitted to a standard protein blast at http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi in a Non-redundant protein sequences (nr) database with a protein-protein blast (blastp) algorithm with expect threshold of 20000 and with PAM 30 matrix and then validated.

The second search approach was in *Homo sapiens* database with none enzyme selection, 40 ppm MS tolerance for precursor ions, 0.5 Da tolerance for fragment ions, +1 peptide charge and with Phospho (ST), Phospho (Y), HexNac(N) and Glu-> pyro Glu (N-term-E) as variable modifications.

Identified peptides were discarded if the unique peptide score were below the minimal Mascot score comparing with the sample score. Modified protein abundance indexes (emPAI) were retrieved for each of the peptide for further quantifications and measurements.

Peptides found in each sample were associated with the proteins that they belong and a list of peptides and proteins was created according with the different groups established before. Protein-protein interaction analysis was performed with Cytoscape v3.1.1 and with ClueGo v2.1.3 (Biological process database 22-09-2014) and CluePedia.

Terminal-N and Terminal-C residues of each studied individual were analysed by type of cancer and type of group and the most common amino acids cleaved residues were submitted to Merops peptidase database (http://merops.sanger.ac.uk/cgi-bin/specsearch.pl) and tables of possible peptidases by amino acid residues were created.

12. RESULTS

12.1 First search approach: Virus:

To identify viral peptide/proteins presents in saliva we preformed the first search approach as described before to all of the subjects in Table VI. We were able to find 29 different viral peptides from 27 different virus species as listed in Table IX.

From the 54 studied individuals only 18 cancer patients presented viral peptides associated with different types of virus as listed in the Table IX. None of the control subjects presented any viral peptide/ proteins.

Results from this method were compared with a previous study from the same research group with the same patients in which was tested the prevalence of HPV, and EBV in saliva and tissue samples of head and neck cancer patients.

68 patients (65 male and 3 female) with age ranges between 34 and 87 were included in the study. HPV was detected in DNA saliva and tissues by restrict fragment length polymorphism (RFLP) using MY09/11 and confirmed by GP5+/6+ primers protocols. Detection of EBV was performed using an in house real time-PCR protocol. The table VII and VIII shows the results of that study.

Table VII - HPV results from the same studied individuals by standardized methods from a previous study from the same research group published in the University of Aveiro Repository by Esteves and all with the title "Infeção por HPV e EBV em tumores de cabeça e pescoço "

HPV	**Saliva (n=66)**		**Biopsies (n=42)**	
	Positive	**Negative**	**Positive**	**Negative**
Male	1	62	2	38
Female	1	2	1	1

Table VIII - EBV results from the same sudied individuals by RT-PCR from a previous study from the same research group published in the University of Aveiro Repository by Esteves and all with the title "Infeção por HPV e EBV em tumores de cabeça e pescoço "

EBV	**Saliva (n=66)**		**Biopsies (n=42)**	
	Positive	**Negative**	**Positive**	**Negative**
Male	46	17	22	18
Female	3	-	-	2

Table IX - Viral peptides, proteins and virus associated found in the studied subjects. None of the controls exhibit any viral peptides.

Patient	Peptides	Start peptide	End Peptide	Score	Protein	Species
4	AHLEPSQRGKKRKRVDDDAGGSAPAQHVPPPQLDHPGREAI	52	92	152.17	EBNA3_EBVA8	*Epstein-Barr virus (human herpesvirus 4)*
13	YDVLPPELWVK	2	12	153.88	YR901_MIMIV	*Acanthamoeba polyphaga mimivirus*
17	VMDAADCFRNSADR	574	587	22.60	RDRP_BTV	*Bluetongue virus*
18	CGCQDKHHDRHDPCV	49	63	99.45	YR623_MIMIV	*Acanthamoeba polyphaga mimivirus*
19	ENTESYAGPDRTAVVGGGFLTTVDQSS	22	48	152.69	POLG_AEVCA	*Avian encephalomyelitis virus*
	PTAAAPRPRPPPPGVGPGGGADPSHPPSRPFRLPPRLALRLRV	96	138	152.69	ICP34_HHV11	*Human herpesvirus 1*
	GASLAGKAAKAACSVVDVYAPSFEPYLH	2998	3025	111.35	RPOA_EAVBU	*Equine arteritis virus*
	LQHNGCIAWEFDDEKESPGQGSCYFYTNSH	118	147	107.68	EV164_ASFK5	*African swine fever virus*
	PPDNPPPPRRARRSQRHADARGPPP	723	747	91.67	POLN_RUBVB POLN_RUBVC	*Rubella virus*
20	VPPPPNPSR	212	220	162,25	PHOSP_MEASC V_MEASC	*Measles Virus*
	PPVNDLKY	880	887	152,09	POLG_DEN4T	*Dengue virus*

					POLG_DEN4P POLG_DEN4S POLG_DEN4H	
21	PGAQRAGGPPSTCSPSSPTSWSCCSSSPRRTRGWGPAGGPVRDS	2639	2682	88.78	TEGU_EHV2	*Equid herpesvirus 2*
	CLPTLVDCIRNCISKVLGYTVIAMPEIGDEEETVQM	803	838	50,32	ENV_FIVU8	*Feline immunodeficiency virus*
22	PRPAGATGGGGGGVHANGGSVFGHPTGGTSTPAHPPP	189 194	225 230	47,04	LT_POVMC LT_POVMA LT_POVM3	*Murine polyomavirus*
23	PPPPPVPQED	711 709	720 718	100.85	UL69_HCMVA UL69_HCMVM	*Human cytomegalovirus (Human herpesvirus 5)*
	PPPPPIRSPD	648	657	100.85	POLR_TYMV POLR_TYMVC	*Turnip yellow mosaic virus*
	QESECVCVNGSCFTIMTDGPSNG	227	249	133,38	NRAM_I00A1 NRAM_I80AA NRAM_I83A1	*Influenza A virus*
24	PPVTPAPPTRAP	35	46	97.12	GL_SUHVK	*Suid herpesvirus 1*
28	PPVTPAPPTRAP	35	46	134.64	GL_SUHVK	*Suid herpesvirus 1*
	PQGPLPAPPNSKTKSMFKRPGRGSVRSLK	440	468	49.75	ATI2_EHV4	*Equid herpesvirus 4*
30	YVGNGTWVHNNTFNVTRYD	39	57	66.55	GH_EHV4	*Equid herpesvirus 4*

31	PPDQASD	2368	2374	152.06	POLG_HCVJA POLG_HCVR6	*Hepatitis C virus*
32	PRLPGPNPVSIERKDFEKLKQ	54	74	152.80	MCE_PBCV1	*Paramecium bursaria Chlorella virus 1*
	LLLVVDFVFVIILLLVLTFVVPRLQ	9	33	103.83	MVP2_GLRV3	*Grapevine leafroll-associated virus 3*
	QPGADGFELQHPY	221	233	52.45	POLG_HE71B POLG_HE71M	*Human enterovirus 71*
36	PPSPPAPAAAPRPSASSASATSSSAAASPAPAPEPARPPRRKRRS	425	469	128.92	IE18_SUHVF	*Suid herpesvirus 1*
37	PGGPYKVGTSGSCPNVTNGNGFFATMAWAVPKNDNNKTA	131 146	169 184	62.53	HEMA_INBNA HEMA_INBF2 HEMA_INBBE HEMA_INBVK	*Influenza B virus*
40	PPPPPVPQED	711 709	720 718	70,56	UL69_HCMVA UL69_HCMVM	*Human cytomegalovirus (Human herpesvirus 5)*
	PPPPPIRSPD	648	657	70,56	POLR_TYMV POLR_TYMVC	*Turnip yellow mosaic virus*
41	PGAPGVKGFGFLNGDNTWLGRTISPRSRSGFEMLKIPNAGTD	344	385	31.44	NRAM_I85A3	*Influenza A virus*
43	TLLGNKTFRTSFVDGAVLETNGPERHN	565	591	152.28	POLN_HEVBU POLN_HEVCH	*Hepatitis E virus*

				POLN_HEVHY	
				POLN_HEVPA	
PVNLLGRSVLQSIVTKFTLAAHTKQIQPLPV	517	547	152.28	POL_JEMBR	*Jembrana disease virus*
LISDIGLEEDNGAID	344	358	152.28	V120_GAHVM	*Gallid herpesvirus 2*

12.2 Second search approach: *Homo sapiens* modifications:

Protein-Protein Analysis

To identify human peptide/proteins presented in saliva we preformed the second search approach as described before to all of the subjects. Clinical characteristics of the subjects enrolled in this study are presented in Table VI**:** head and neck cancer patients and 10 healthy control subjects. After submitted the samples to nanoHPLC separation and MS/MS spectra analysis peptides and proteins were identified with Mascot software and emPAI values were collected for each subject. 123 salivary proteins were identified and divided according to the different groups established before. Figure 7 illustrates the distribution of identified proteins per group. More detailed schemes can be found in the supplementary figures.

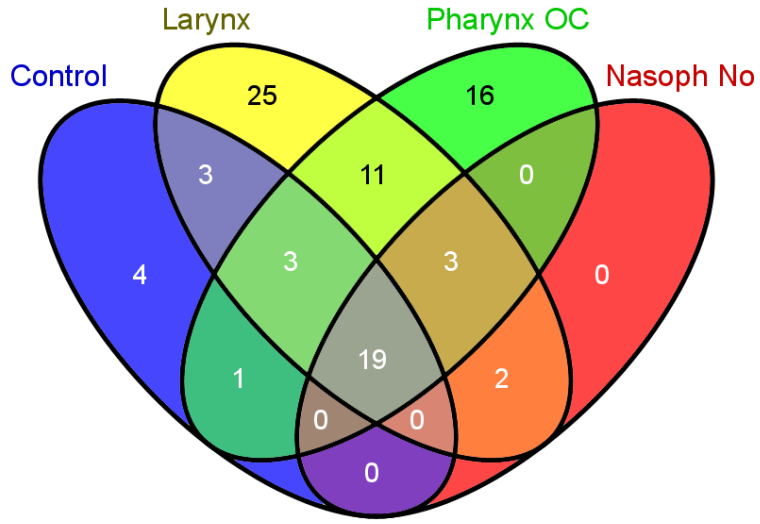


Figure 7 - Protein distribution by control and cancer groups. Analysis preformed with Venny

From all of the salivary proteins detected only 19 were common to all of studied subjects including proline rich-proteins, histatin 1, mucin 7, SMR3B protein, statherins, cathepsins, H2A2A and H4 histone, cytoplasmic actin, fibrinogen beta chain, hemoglobin, lactotransferrin and the S100A9 protein. These common proteins are mainly involved in the salivary processes specified in Table X**.**

Table X - Most abundant proteins in studied saliva samples and they're biological function

Protein	Biological Function
ACTB	Platelet aggregation, Retina Homeostasis
Cathepsin G	Selection defence response to fungus, killing of cells of other organism, disruption of cells of other organism
Fibrinogen beta chain	Platelet aggregation, homotypic cell-cell adhesion
Hemoglobin subunit beta	
Histatin 1	Defence response to fungus, killing of cells of other organism, disruption of cells of other organism
Lactotransferrin	Regulation of lipopolysaccharide-mediated signalling pathway, lipopolysaccharide-mediated signalling pathway
Basic salivary proline-rich protein 3	Defence response to bacterium
S100A9	Regulation of lipopolysaccharide-mediated signalling pathway, defence response to fungus, inflammatory processes
Statherin	Regulation of bone mineralization and biomineral tissue development
Zymogen granule protein 16 homolog B	Retina homeostasis (carbohydrate binding)

From the distribution scheme from figure 7 unique proteins to each group were identified, protein-protein interaction was analysed and we correlate several cellular pathways with saliva proteome.

It was possible to identify 4 proteins exclusive to Control Group (Blue), 25 proteins exclusive to Larynx group (yellow), 16 proteins exclusive to Pharynx and Oral Cavity and none exclusive to Nasopharynx and Nostrils Group.

Protein-protein interaction was preformed with Cytoscape and Go Clue and 5 main clusters were identified. In figure 8 it is possible to identify five clusters

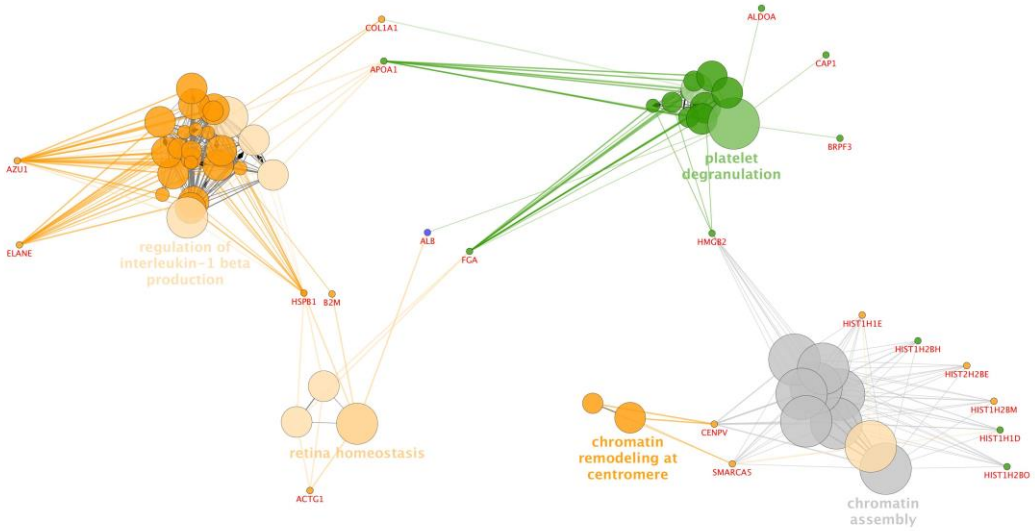


Figure 8 - Protein- protein interaction from the unique proteins found in figure 7 were preformed with Cytoscape and CluePedia. Yellow clusters belong mainly to larynx group and green cluster belong mainly to pharynx and oral cavity group. Yellow dots are proteins from the larynx group, green dots are proteins from Pharynx and Oral Cavity group and Blue dots are proteins from Control group

with biological functions like retina homeostasis, platelet degranulation, regulation of interleukin – 1 beta production, chromatin assembly and chromatin remodelling ate centromere.

The cluster in green has genes that encode proteins mainly from Pharynx and Oral Cavity group (green) and with one gene from Control group (blue). This cluster is composed by 8 genes from which 5 share functions with other clusters. Genes like

ALDOA, BRPF3 and CAP1 encode proteins that have mainly functions in platelet degranulation; shared genes with retina homeostasis cluster like ALB and FGA also encode proteins that regulates heterotypic cell-cell adhesion and negative regulation of extrinsic apoptotic signalling pathway via death domain receptors. APOA1, COL1A1 and HMGB2 share functions with regulation of interleukin-1 beta production cluster and have biological functions in positive regulation of substrate adhesion-dependent cell spreading, positive regulation of cell morphogenesis involved in differentiation, regulation in collagen catabolic process, negative regulation of extrinsic apoptotic signalling pathway via death domain receptors and also in regulation of heterotypic cell-cell adhesion.

HMGB2 has also chemo attractant activity binding proteins associated with chromatin and single stranded DNA and it also has the ability to bend DNA.

Additionally is possible to identify several yellow clusters mainly from the Larynx group (yellow). The biggest cluster is constituted by six genes mainly with the biological function of regulation of interleukin – 1 beta production.

From the six genes, APOA1 and COL1A1 have shared functions with platelet degranulation cluster and B2M and HSPB1 with retina homeostasis cluster. Besides the retina homeostasis functions, B2M and HSPB1 proteins also play roles in leukocyte mediated cytotoxicity and positive regulation of tumour necrosis factor biosynthetic process.

AZU and ELANE limited to this cluster have biological functions in the positive regulation of tumour necrosis factor biosynthetic process and in regulation of chemokine biosynthetic process.

Three genes compose the smaller yellow cluster, the retina homeostasis cluster: B2M, ACTG1, HSPB1 (described before) with platelet aggregation functions.

The last cluster is composed by eight genes that encodes proteins from Larynx group and Pharynx and Oral Cavity group and one shared gene with platelet degranulation cluster from Pharynx and Oral Cavity group proteins and have main biological functions in chromatin assembly and chromatin remodelling at centromere - Figure 9.

From the eight exclusive gens to this cluster HIST1H1D (Histone H1.3), HIST1H2BH (Histone H2B type 1-H) and HIST1H2BO (Histone H2B type 1-O)

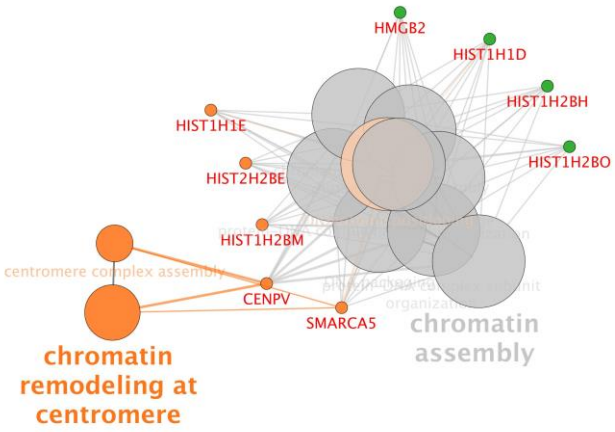


Figure 9 - Cluster with proteins from Larynx group (yellow) and Pharynx and Oral Cavity group (green)

belong to Pharynx and Oral Cavity group and HIST1H1E (Histone H1.4), HIST2H2BE, HIST1H2BM, CENPV (Centromere protein V) and SMARCA5 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5) belong to Larynx group.

HIST1H1D encodes for proteins with functions in nucleosome positioning and assembling by forming the chromatin fibber regulating the transcription. Post translational modifications in Histones H1 usually are phosphorylations during phases G2 and M in cell cycle and dephosphorylated sharply after.

HIST1H2BH and HIST1H2BO is a core component of the nucleosome, playing a central role trough chromatin organization in transcription regulation, DNA repair, DNA replication and chromosomal stability.

From the Larynx group CENPV is a protein required in the centromeric formation and organization once is responsible for the distribution pericentromeric heterochromatin in interphase, in mitosis.

SMARCA5 is a helicase that processes ATP-dependent nucleosome-remodelling activity. Complexes with this gene are capable of forming ordered nucleosome arrays on chromatin which requires H4 histones and are required in S phase for replication of pericentric heterochromatin and is a an essential component of the WICH complex that regulates the transcription of various genes and mediates the histone H2AX phosphorylation.

HIST2H2BE besides the DNA binding activity similar with the histones described before has a broad antibacterial activity especially against Gram-Positive bacteria's creating an innate immune response in mucosa. HIST1H1E and HIST1H2BM similar with HIST1H2BO is also a core component of the nucleosome, which will warp and compact the DNA into chromatin, it will also have a central role in the transcription regulation, DNA repair, DNA replication and chromosomal stability. All of these proteins from this cluster are correlated with the cell cycle formation.

Peptide Analysis

Sequences of the peptides were also studied, a total of 1885 peptide sequences were identified and 104 sequences were found to be common to the four different groups, 168 sequences were found only in Control group, 436 only in Larynx group, 430 only in Pharynx and Oral Cavity group, 59 in Nasopharynx and Nostrils Group and 14 sequences were common to all of the cancer groups – Figure 10.

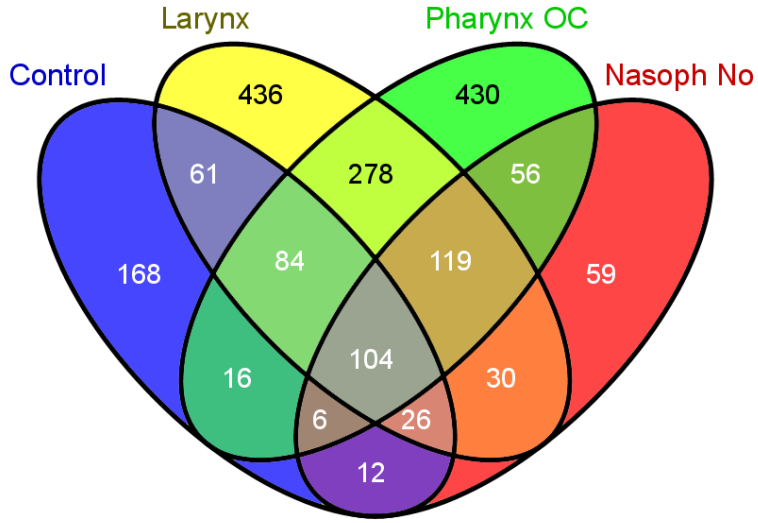


Figure 10 – Peptide distribution by group. Analysis was performed with resource to Venny

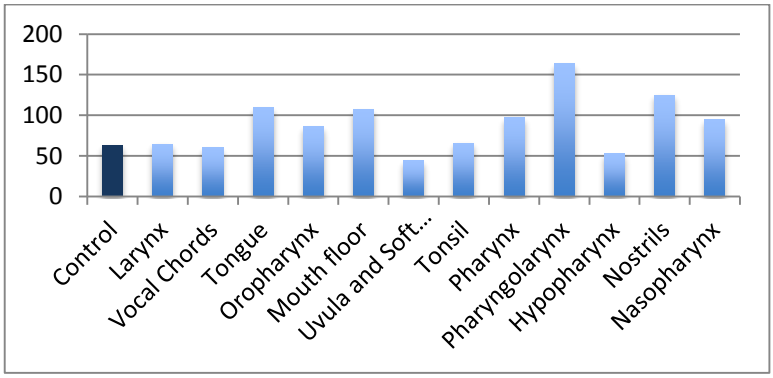
Sequences common to all of the groups predominantly belong to salivary proteins like proline rich-proteins, histatin 1, mucin 7, SMR3B protein, statherins, cathepsins, H2 and H4 histone, actin, fibrinogen beta chain and hemoglobin.

It was possible to identify 168 peptides exclusive to Control Group (Blue), 436 peptides exclusive to Larynx group (yellow), 430 peptides exclusive to Pharynx and Oral Cavity and 59 peptides exclusive to Nasopharynx and Nostrils Group.

In order to evaluate the proteolitic activity between control and all types of cancers the number peptide sequences of each studied individual were counted and

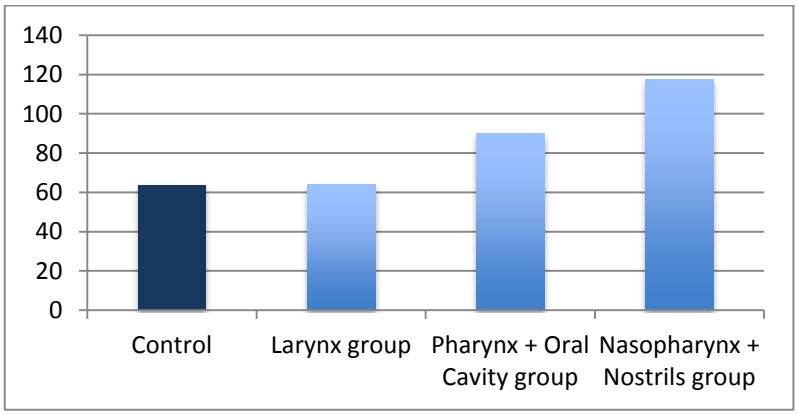
associated by type of cancer and by cancer groups established before. In Graphic 1 we can observe the average number of unique sequences for each type of cancer and for controls:

Graphic 1 - Comparation of the distribution of the number of unique sequences by cancer against control



In Graphic 2 we can observe the average number of unique sequences for cancer groups established before and for controls:

Graphic 2 - Comparation of the distribution of unique sequences by the groups established before.

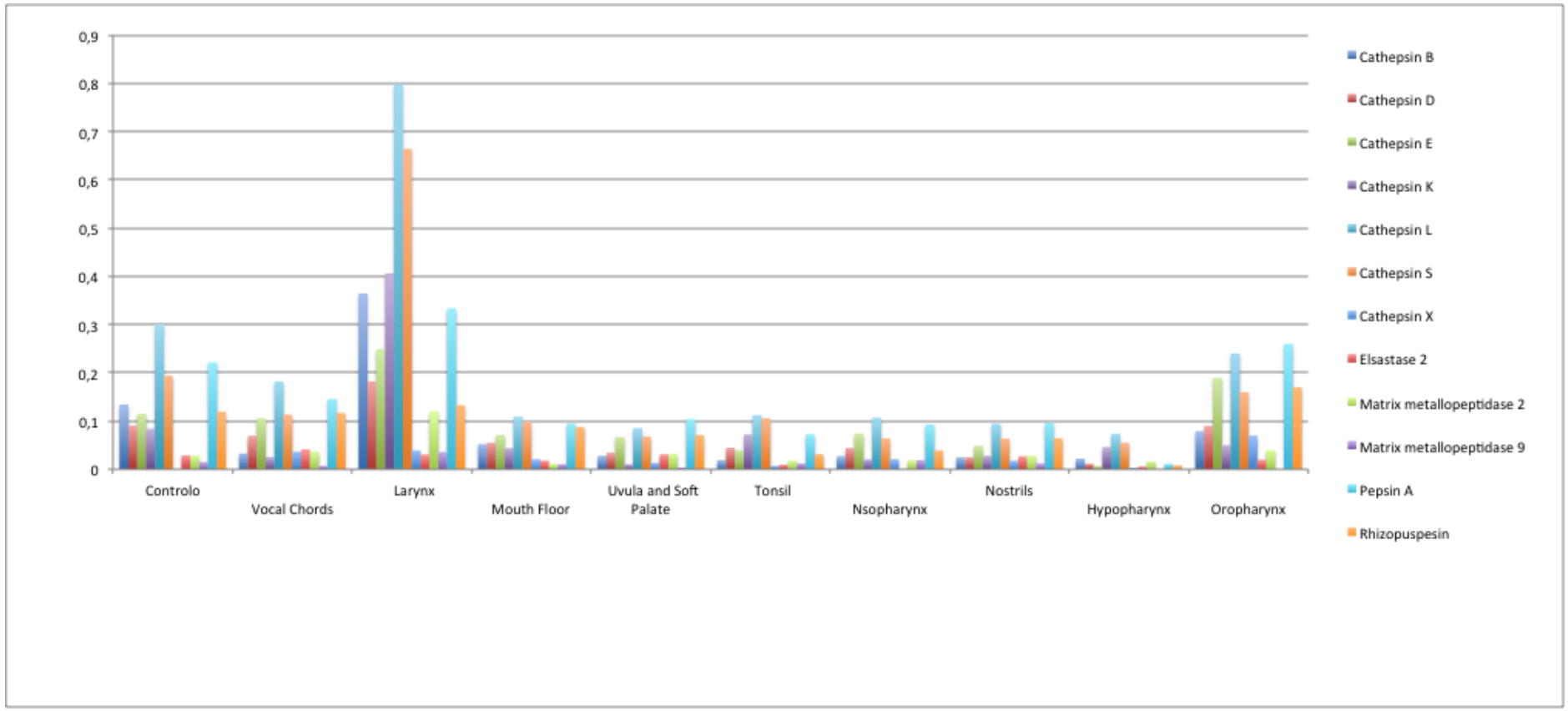


To each type of cancer aminoacids of the terminal-N and terminal-C of the peptide residues were selected by the highest number of clivage repeats. In Merops peptidase database, terminal-N and terminal-C residues were searched and tables with the possible salivary proteases were created in order to predict the proteolitic events in saliva of our patients (supplementary tables).

In order to normalize the number of cleavages for each type of cancer a clivage ratio was created dividing the number of cleavages for the number of peptide sequences of all samples. Then in each type of cancer for each type of peptidase, frequency values were created and compared by each type of cancer – Graphic 3.

Peptidases found were compared between cancers and controls and we could observe that the most abundant peptidases in our samples are cathepsins B, S and L and Pepsin A.

Graphic 3 - Peptidase distribution by type of cancer



Comparing the peptidases found by cancer site against controls it was possible to observe that larynx and oropharynx have a higher proteolitic activity than controls. We can also observe that Cathepsin X wasn’t present in the control group and was present in all cancer types.

The rest of the cancers doesn’t exhibit distinguish values from the control group.

13. Discussion

13.1 First search approach: Virus:

As we have seen before cancer evolve through a series of somatic mutations that will induce cytogenetic changes resulting in uncontrolled cellular proliferation(111, 112). According to IARC there are several risk factors like alcohol, tobacco or virus that can increase the probability of breeding a cancer.

In order to be possible to associate a specific to human cancer in 1990 epidemiologic and virologic guidelines were established, this way it is possible to relate a cancer to a virus if the prevention of infection (vaccination) or the delay of the time infection decreases the incidence of the tumour; if the virus in vitro is able to transform normal cells into malignant cells; if the virus genome or DNA is demonstrated only in tumour cells and not in normal cells and if the virus is able to induce tumour in experimental animals and if its neutralization prevents the tumour development(113).

Several studies indicate that the most frequently associated virus to oral cancer progression and transformation have been Herpes group virus, HPV (114, 115), adenoviruses (116) and hepatitis C (117, 118). Despite human oncogenic viruses exploit different strategies that will contribute to cancer progression and development, they all share the propensity to establish long-term persistent infections using evading strategies to avoid the host immune response(113).

In our study it was possible to identify several viral peptides in saliva from different types of viruses (including animal viruses), however none of the control subjects presented any type of viral peptide, which fulfil one of the guidelines established before. In table XI it is possible to prevent a strength of association between the viruses found and the patient cancers according with published researches.

Virus	Tumour	Association with literature
Acanthamoeba polyphaga mimivirus	Laryngeal SCC	Related with pneumonia infections; no association found with cancers (119-121) *
Bluetongue virus	Laryngeal SCC	Predominant in ruminants; Several studies in human cancer cells suggests this virus have cytotoxic effects and can induce apoptosis (122-124) *
Avian encephalomyelitis virus	Laryngeal SCC	Related to hepatitis A; possesses motifs characteristic of H-rev107 proteins involved in cell proliferation; modulate apoptosis (125-127) *
Gallid herpesvirus 2 (Avian herpesvirus 1)	Nostrils	oncogenic herpesvirus in chickens that causes lymphoma; like human lymphoma chicken lymphoma cells overexpress the CD30 antigen; similar Epstein-Barr and herpesvirus-6; interacts directly with p53 and inhibits p53-mediated transcriptional activity and apoptosis (128-131) *
Jembrana disease virus	Nostrils	similar to bovine immunodeficiency virus; owns genes in common with HIV (132, 133) *
Suid herpesvirus 1	Laryngeal SCC Pharyngeal SCC Oropharyngeal SCC	Similar to herpesvirus, there aren't any studies in human cellular lines
African swine fever virus	Laryngeal SCC	The ORF A179L of the virus is similar to proto-oncogene bcl-2; and in HeLa (human), myeloid leukaemia cell line K562 (human) and BSC-40 (monkeys) cell lines protects cells from programmed cell death (134, 135) *

Equine arteritis virus	Laryngeal SCC	causes respiratory and reproductive disease in horses; induces apoptosis initiated by caspase-8 activation and mitochondria-dependent caspase-9 activation; despite infect HeLa cells persistently there are no studies relating the virus with cancers (136-138) *
Equid herpesvirus 2 (equine cytomegalovirus)	Laryngeal SCC	Differs substantially from equine herpesvirus 1 and 3; induces oncogenic transformation in hamster embryo cells (139, 140) *
Equid herpesvirus 4	Mouth floor SCC Oropharyngeal SCC	Induces equine respiratory diseases; similar to equine herpesvirus 1; no association found with cancers (141)
Murine polyomavirus	Laryngeal SCC	Has oncogenic properties in murines; belongs to the family of polyomavirus; human polyomavirus induces tumours (142, 143) *
Feline immunodeficiency virus	Laryngeal SCC	Causes immune deficiency in cats; can infect humans but enters in latent stage; it's a possible vector that can be used against human cancers (144, 145) *
Grapevine leafroll-associated virus 3	Uvula and Soft palate SCC	Transmitted by bugs to grapevines; no association found with cancers (146, 147)
Turnip yellow mosaic virus	Laryngeal SCC Hypopharyngeal SCC	Infects green vegetables; no association found with cancers (148)
Paramecium bursaria	Uvula and Soft	Infects green algae; no association

Chlorella virus 1	palate SCC	found with cancers (149, 150)
Human enterovirus 71	Uvula and Soft palate SCC	Pathogen of hand-foot-and-mouth disease; associated with severe neurological disease; no association found with cancers (151)
Measles Virus (rubeola)	Laryngeal SCC	Causes contagious respiratory infection; no association found with cancers (152)
Rubella virus	Laryngeal SCC	Similar to measles; provoke severe rashes; no association found with cancers (152)
Dengue virus	Laryngeal SCC	Mosquito-borne disease; lead to massive bleeding, shock, and death; doesn't increase tumour necrosis factor-α-induced permeability; no association found with cancers (153)
Influenza A virus	Laryngeal SCC Nostrils	Flu with respiratory complications; no association found directly with cancer (154, 155)
Influenza B virus	Pharyngolaryngeal SCC	Similar to influenza A (154, 155)
Hepatitis C virus	Mouth floor SCC	Type 1 carcinogenic agent by IARC; oncogenic virus Induces hepatocellular cancer;
Human herpesvirus 1	Laryngeal SCC	Related to several cancers; not related to cancer Laryngeal cancer
Epstein-Barr virus (human herpesvirus 4)	Vocal Chords SCC	Related to nasopharyngeal Carcinoma; EBNA 3 (the protein found) is essential for B-cells

			transformation and EBV immortalization; overtake the pRb check points in the cell cycle (34)
	Human cytomegalovirus (Human herpesvirus 5)	Laryngeal SCC Hypopharyngeal SCC	Infection arises in cancer patients due to compromised immunitary system

Table XI - Viruses associated with viral peptides and association with current literature

Relating a virus to cancer is a complicated process with several difficulties associated. The fact of the human host to be unique, immune system status, single genetic characteristics at molecular level and the age at the time of the infection turn this multistep progress disease hard to follow and associate to virus. Also causes of cancer may vary by geographical areas and most related cancers need co-factors for progression(113).

Other factors to have in consideration are the different oncogenic potential in different strains (ex. HPV), the long incubation time between infection and the cancer and the inability to reproduce many human cancers with the virus in study in experimental animals(113).

As it was possible to verify in our state of the art, there are several viruses that have the capability to induce the lost of control in normal cells and change them into tumoural cells. In addiction, we also saw that several viral infections like HIV can diminish the immunitary system functions and its capacity to detect and destroy possible carcinogenic cells.

In our study we didn't find any viral peptides in the control group, which fulfil one of the guidelines to relate a virus to a cancer, this lead us to think that probably even the animal viruses that we found could contribute in any unknown way to

mutagenesis or could compromise the normal immune system function allowing a easier cancer development.

Despite we couldn't correlate the viruses with the specific cancers sites of our patients it lead us to think if there are animal viruses that can be starting to induce changes in human cells once some of them have the capability to modulate apoptosis protecting the cell from a programed death.

We found several animal viruses with similar proteins to human viruses and human proto-oncogenes like *Avian encephalomyelitis virus* similar to *Hepatitis A*, *Jembrana disease virus* similar to *Human immunodeficiency virus* or the ORF A179L from *African swine fever virus* similar to bcl-2

Animal virus should not be able to disarray human cells once they enter in human system and usually stay in latency state, not causing any signs of symptoms of the animal disease in humans. However viruses usually suffer several mutations evolving and unknown strains are constantly being created.

In the case of the *Gallid herpesvirus 2 (Avian herpesvirus 1),* an oncogenic viruses causing nasal cancer in chickens and with similar proprieties to EBV, known to provoke nasopharyngeal carcinomas in humans, the cancer development site is the same in the chickens and in the humans. We couldn't found more studies that could relate the virus specifically to human cancer.

Still we also found viruses like the *Feline immunodeficiency virus* that due to its similarity to human viruses are starting to be manipulated and used like vectors in new cancer researches to infect the cancer cells and destroy them.

This is the first study relating viral peptides and saliva from head and neck cancer patients until today and we expect that will come to contribute to new non-invasive methods of prognosis or risk trough saliva that can measure viruses as biomarkers.

13.2 Second search approach: *Homo sapiens* modifications:

Our study compares the salivary proteome between controls and head and neck cancer patient, allowing the foreseeing of potential disease biomarkers with diagnosis and prognosis significance and the future elucidation of disease pathogenesis.

From the 123 proteins identified in the saliva of all subjects studied only 3 proteins were exclusively assigned to cancer patients. Analysis with STRING bioinformatics tool suggests their involvement in condensation of nucleosome chains into higher order structures, cytoskeleton organization and oxygen transportation to peripheral tissues. Actually clinical studies in prostate and breast relate the involvement of Histones 1 (Histone H1.5) as a possible diagnostic biomarker(156, 157).

The analysis of salivary proteins unique by group revealed four distinct clusters. Mainly composed by proteins from pharynx and oral cavity group we have found the platelet degranulation cluster, which is a biological process related to blood homeostasis with association to malignant diseases since the 19th century (158). Platelets are activated when an injury occurs to prevent fatal bleedings; a growing tumour is dependent on angiogenesis for oxygen and nutrient supply as well as for removal of waste products. Platelets are also metastasis regulators (158).

Mainly composed by proteins from the Larynx group we have found a cluster related with regulation of Interleukin-1 beta production. Interleukins are cytokines with central roles in regulation of immune and inflammatory responses, several studies show different opinions on whether this cytokine is related to oral cancer or not. A study from 2007 studied the +3953 C/T polymorphism and conclude that interleukin-1 beta doesn’t play a primary role in this malignancy since there are other interleukins with more prominent effects(159). However a study from 2013 preformed in unstimulated whole saliva in OSCC patients pre and pos-operation detected different cytokine concentration concluding that might be a useful tool in early stages detection(160).

Another significant cluster it was a cluster composed mainly with proteins from larynx and pharynx and oral cavity groups related to chromatin functions composed mainly by Histones 1 and Histones 2.

The fundamental unit of chromatin, the nucleosome is composed by an octamer of histones from types H2A, H2B, H3 and H4, this structure affects primarily the wrapping of DNA and then the replication, transcription and the enzymes involved in the process(161). Monteiro et all relate different types of cancer with different types of H2 histones Chen et all was able to associate several Histones 3 to different prognosis in Oral squamous cell carcinoma(162). Histones can also suffer post-translational mutations that can alter the chromatin structure and gene transcription originating several variants. The variants so far associated to oral cancer are H3.3 and H2A.X (163). In histones cluster is it possible to identify one protein – SMARCA5 – one essential component to the WICH complex that regulates H"AX phosphorylation which lead us to believe that this cluster might have histones that can be used in the future as biomarkers to our cancers. One of our patients presented the protein to H2A.X however since it was the only one and our analysis are based in the proteins present in at least two individuals that result could not be verified and wasn't considered in the analysis.

In our study we also preformed the peptides analysis in order to disclosure peptide with more abundance and the potential associated salivary proteases to head and neck cancers. We intend to see if the cancer patients suffer from more proteolitic activity than the control patients since in our state of the art we have found some studies relating the rate of proteolysis with certain diseases.

It was possible to identify that most of the common sequences of our patients belong to proteins with salivary functions.

Most of the types of cancer present a higher number of unique sequences than controls, the cancer type that we have found with more unique sequences was the pharyngolarynx cancer, however, we must take under consideration that our cancers have different numbers of patients which are unique and once one person can present more peptides and proteins in saliva than other.

Proteolytic analysis revealed several proteases, the most common were: Cathepsins (B, D, E, K, L and X), Elastase-2, Matrix metallopeptidases 2 and 9, Rhizopuspesin and pepsin A.

Comparing each cancer to the control group we have found that larynx and oropharynx samples suffers more cleavage than the control group.

A proteolytic analysis from Schmidt et all in tissue samples from OSCC patients revealed matrix metallopeptidase-9 increased the gelatinolytic and fibrinolytic activity in tumour samples (164). In our samples for matrix metallopeptidase -9 the cleavage ratio was higher than controls in Nasopharynx and Larynx cancer.

This analysis could indicate a possible biomarker by cleave rate for these specific cancers.

14. CONCLUDING REMARKS

This work addresses the first exploratory study regarding salivary proteases and peptidome with head and neck cancer patients. Results lead us to think in each way animal viruses are implicated in some human cancers and if they're latency state help in the development of cancer.

Proteolytic analysis revealed that maybe it would be possible to detect larynx and oropharynx cancers in previous stages but we need further studies to comprise that.

1. BIBLIOGRAPHY

1. Dufour X, Beby-Defaux A, Agius G, Lacau St Guily J. HPV and head and neck cancer. European Annals of Otorhinolaryngology, Head and Neck Diseases. 2012;129(1):26-31.
2. Kufe DW, Frei E, Holland JF, Pollock RE, Bast RC, Weichselbaum RR. Holland-Frei's Cancer Medicine Review: Decker Publishing; 2003.
3. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
4. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011;144(5):646-74.
5. Goldman L, Schafer AI. Cecil Medicine: Expert Consult - Online: Elsevier Health Sciences; 2011.
6. Montazeri A. Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008. Health and Quality of Life Outcomes. 2009;7:102-.
7. Whale Z. Head and neck cancer: an overview of literature. European Journal of Oncology Nursing. 1998;2(2):99-105.
8. Ferlay J SH, Bray F, Forman D, Mathers C and Parkin DM. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet] Lyon, France: Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]; 2008 [cited 2010 2012]. Available from: http://globocan.iarc.fr.
9. Crozier E, Sumer BD. Head and neck cancer. Medical Clinics of North America. 2010;94(5):1031-46.
10. Cancer IAfRi. Agents Classified by IARC Lyon, France2013 [cited 2013 2013]. Available from: http://monographs.iarc.fr/ENG/Classification/index.php.
11. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. The Lancet.371(9625):1695-709.
12. Petti S. Lifestyle risk factors for oral cancer. Oral Oncology. 2009;45(4-5):340-50.
13. Gondivkar SM, Parikh RV, Gadbail AR, Solanke V, Chole R, Mankar M, et al. Involvement of viral factors with head and neck cancers. Oral Oncology. 2012;48(3):195-9.
14. Yu GP, Zhang ZF, Hsu TC, Spitz MR, Schantz SP. Family history of cancer, mutagen sensitivity, and increased risk of head and neck cancer. Cancer Letters. 1999;146(1):93-101.
15. Scully C, Field JK, Tanzawa H. Genetic aberrations in oral or head and neck squamous cell carcinoma (SCCHN): 1. Carcinogen metabolism, DNA repair and cell cycle control. Oral Oncology. 2000;36(3):256-63.
16. Pelucchi C, Tramacere I, Boffetta P, Negri E, Vecchia CL. Alcohol consumption and cancer risk. Nutrition and Cancer. 2011;63(7):983-90.
17. Hardt M, Thomas LR, Dixon SE, Newport G, Agabian N, Prakobphol A, et al. Toward defining the human parotid gland salivary proteome and peptidome: Identification and characterization using 2D SDS-PAGE, ultrafiltration, HPLC, and mass spectrometry. Biochemistry. 2005;44(8):2885-99.
18. Ha PK, Chang SS, Glazer CA, Califano JA, Sidransky D. Molecular techniques and genetic alterations in head and neck cancer. Oral Oncology. 2009;45(4–5):335-9.

19. Uchida K, Oga A, Okafuji M, Mihara M, Kawauchi S, Furuya T, et al. Molecular cytogenetic analysis of oral squamous cell carcinomas by comparative genomic hybridization, spectral karyotyping, and fluorescence in situ hybridization. Cancer Genetics and Cytogenetics. 2006;167(2):109-16.
20. Singh B, Gogineni SK, Sacks PG, Shaha AR, Shah JP, Stoffel A, et al. Molecular cytogenetic characterization of head and neck squamous cell carcinoma and refinement of 3q amplification. Cancer Research. 2001;61(11):4506-13.
21. Patmore HS, Ashman JNE, Stafford ND, Berrieman HK, MacDonald A, Greenman J, et al. Genetic analysis of head and neck squamous cell carcinoma using comparative genomic hybridisation identifies specific aberrations associated with laryngeal origin. Cancer Letters. 2007;258(1):55-62.
22. You L, He B, Xu Z, McCormick F, Jablons DM. Future Directions: Oncolytic Viruses. Clinical Lung Cancer. 2004;5(4):226-30.
23. Ewald PW, Ewald HAS. Infection, mutation, and cancer evolution. Journal of Molecular Medicine. 2012;90(5):535-41.
24. Young LS, Murray PG. Epstein-Barr virus and oncogenesis: from latent genes to tumours. Oncogene. 2003;22(33):5108-21.
25. Nowsheen S, Aziz K, Panayiotidis MI, Georgakilas AG. Molecular markers for cancer prognosis and treatment: Have we struck gold? Cancer Letters. (0).
26. Zur Hausen H. Papillomavirus infections - A major cause of human cancers. Biochimica et Biophysica Acta - Reviews on Cancer. 1996;1288(2):F55-F78.
27. De Villiers EM, Fauquet C, Broker TR, Bernard HU, Zur Hausen H. Classification of papillomaviruses. Virology. 2004;324(1):17-27.
28. De Villiers EM. Human pathogenic papillomavirus types: An update. Current Topics in Microbiology and Immunology. 1994;186:1-12.
29. Nobre RJ, de Almeida LP, Martins TC. Complete genotyping of mucosal human papillomavirus using a restriction fragment length polymorphism analysis and an original typing algorithm. Journal of Clinical Virology. 2008;42(1):13-21.
30. Doorbar J, Ely S, Sterling J, McLean C, Crawford L. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. Nature. 1991;352(6338):824-7.
31. Roberts S, Ashmole I, Johnson GD, Kreider JW, Gallimore PH. Cutaneous and mucosal human papillomavirus E4 proteins form intermediate filament-like structures in epithelial cells. Virology. 1993;197(1):176-87.
32. Doorbar J. The papillomavirus life cycle. Journal of Clinical Virology. 2005;32(SUPPL.):S7-S15.
33. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: a virus-related cancer epidemic. The Lancet Oncology. 2010;11(8):781-9.
34. Thompson MP, Kurzrock R. Epstein-Barr Virus and Cancer. Clinical Cancer Research. 2004;10(3):803-21.
35. Farrell PJ, Cludts I, Stühler A. Epstein-Barr virus genes and cancer cells. Biomedicine and Pharmacotherapy. 1997;51(6-7):258-67.
36. Goldenberg D, Golz A, Netzer A, Rosenblatt E, Rachmiel A, Goldenberg RF, et al. Epstein-Barr virus and cancers of the head and neck. American Journal of Otolaryngology - Head and Neck Medicine and Surgery. 2001;22(3):197-205.

37. Wilkins MR, Sanchez JC, Williams KL, Hochstrasser DF. Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. Electrophoresis. 1996;17(5):830-8.
38. Blackstock WP, Weir MP. Proteomics: Quantitative and physical mapping of cellular proteins. Trends in Biotechnology. 1999;17(3):121-7.
39. Kay R, Barton C, Ratcliffe L, Matharoo-Ball B, Brown P, Roberts J, et al. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. Rapid Communications in Mass Spectrometry. 2008;22(20):3255-60.
40. Montelione GT, Anderson S. Structural genomics: Keystone for a human proteome project. Nature Structural Biology. 1999;6(1):11-2.
41. Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, et al. The human proteome project: Current state and future direction. Molecular and Cellular Proteomics. 2011;10(7).
42. Penque D, Simões T, Amado F. Proteomics advances in the last decade: What is next? Journal of Proteomics. 2011;75(1):1-3.
43. Rodríguez-Ulloa A, Rodríguez R. Bioinformatic tools for proteomic data analysis: An overview. Biotecnologia Aplicada. 2008;25(4):312-9.
44. Chich JF. A mini review: Proteomic analysis, a post-genomic approach. Lait. 2001;81(1-2):13-8.
45. Yates Iii JR, Gilchrist A, Howell KE, Bergeron JJM. Proteomics of organelles and large cellular structures. Nature Reviews Molecular Cell Biology. 2005;6(9):702-14.
46. Messana I, Cabras T, Iavarone F, Vincenzoni F, Urbani A, Castagnola M. Unraveling the different proteomic platforms. Journal of Separation Science. 2013;36(1):128-39.
47. Schulte I, Tammen H, Selle H, Schulz-Knappe P. Peptides in body fluids and tissues as markers of disease. Expert Review of Molecular Diagnostics. 2005;5(2):145-57.
48. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJC, et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. Journal of Proteome Research. 2010;9(10):4982-91.
49. Huang CM. Comparative proteomic analysis of human whole saliva. Archives of Oral Biology. 2004;49(12):951-62.
50. Vitorino R, Lobo MJC, Ferrer-Correira AJ, Dubin JR, Tomer KB, Domingues PM, et al. Identification of human whole saliva protein components using proteomics. Proteomics. 2004;4(4):1109-15.
51. Bonne NJ, Wong DTW. Salivary biomarker development using genomic, proteomic and metabolomic approaches. Genome Medicine. 2012;4(10).
52. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Molecular & cellular proteomics : MCP. 2002;1(11):845-67.
53. Hu S, Loo JA, Wong DT. Human body fluid proteome analysis. Proteomics. 2006;6(23):6326-53.
54. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Prac Oncol. 2008;5(10):588-99.
55. Schaaij-Visser TBM, Brakenhoff RH, Leemans CR, Heck AJR, Slijper M. Protein biomarker discovery for head and neck cancer. Journal of Proteomics. 2010;73(10):1790-803.

56. Dowling P, Wormald R, Meleady P, Henry M, Curran A, Clynes M. Analysis of the saliva proteome from patients with head and neck squamous cell carcinoma reveals differences in abundance levels of proteins associated with tumour progression and metastasis. Journal of Proteomics. 2008;71(2):168-75.
57. Veenstra TD, Prieto DA, Conrads TP. Proteomic patterns for early cancer detection. Drug Discovery Today. 2004;9(20):889-97.
58. Tainsky MA. Genomic and proteomic biomarkers for cancer: A multitude of opportunities. Biochimica et Biophysica Acta - Reviews on Cancer. 2009;1796(2):176-93.
59. Zolg JW, Langen H. How industry is approaching the search for new diagnostic markers and biomarkers. Molecular and Cellular Proteomics. 2004;3(4):345-54.
60. Granger DA, Kivlighan KT, Fortunato C, Harmon AG, Hibel LC, Schwartz EB, et al. Integration of salivary biomarkers into developmental and behaviorally-oriented research: Problems and solutions for collecting specimens. Physiology & Behavior. 2007;92(4):583-90.
61. Castagnola M, Picciotti PM, Messana I, Fanali C, Fiorita A, Cabras T, et al. Potential applications of human saliva as diagnostic fluid. Acta Otorhinolaryngologica Italica. 2011;31(6):347-57.
62. Pfaffe T, Cooper-White J, Beyerlein P, Kostner K, Punyadeera C. Diagnostic potential of saliva: Current state and future applications. Clinical Chemistry. 2011;57(5):675-87.
63. Caseiro A, Vitorino R, Barros AS, Ferreira R, Calheiros-Lobo MJ, Carvalho D, et al. Salivary peptidome in type 1 diabetes mellitus. Biomedical Chromatography. 2012;26(5):571-82.
64. Baldini C, Giusti L, Bazzichi L, Lucacchini A, Bombardieri S. Proteomic analysis of the saliva: A clue for understanding primary from secondary Sjögren's syndrome? Autoimmunity Reviews. 2008;7(3):185-91.
65. Cheng B, Rhodus NL, Williams B, Griffin RJ. Detection of apoptotic cells in whole saliva of patients with oral premalignant and malignant lesions: A preliminary study. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology. 2004;97(4):465-70.
66. Susceptibility to Dental Caries and the Salivary Proline-Rich Proteins. International Journal of Dentistry. 2011;2011.
67. Haigh BJ, Stewart KW, Whelan JRK, Barnett MPG, Smolenski GA, Wheeler TT. Alterations in the salivary proteome associated with periodontitis. Journal of Clinical Periodontology. 2010;37(3):241-7.
68. Gonçalves LDR, Soares MR, Nogueira FCS, Garcia C, Camisasca DR, Domont G, et al. Comparative proteomic analysis of whole saliva from chronic periodontitis patients. Journal of Proteomics. 2010;73(7):1334-41.
69. Zhang A, Sun H, Wang P, Wang X. Salivary proteomics in biomedical research. Clinica Chimica Acta. 2013;415:261-5.
70. Wu Z, Doondeea JB, Gholami AM, Janning MC, Lemeer S, Kramer K, et al. Quantitative chemical proteomics reveals new potential drug targets in head and neck cancer. Molecular and Cellular Proteomics. 2011;10(12).
71. Nagler RM, Ben-Izhak O, Cohen-Kaplan V, Shafat I, Vlodavsky I, Akrish S, et al. Heparanase up-regulation in tongue cancer: Tissue and saliva analysis. Cancer. 2007;110(12):2732-9.

72. Moss SJ. Clinical implications of recent advances in salivary research. J Esthet Dent. 1995;7(5):197-203.
73. Levine MJ. Development of artificial salivas. CRIT REV ORAL BIOL MED. 1993;4(3-4):279-86.
74. Navazesh M. Methods for Collecting Saliva. Annals of the New York Academy of Sciences. 1993;694(1):72-7.
75. Kaufman E, Lamster IB. The Diagnostic Applications of Saliva— A Review. Critical Reviews in Oral Biology & Medicine. 2002;13(2):197-212.
76. Mandel ID, Wotman S. The salivary secretions in health and disease. Oral sciences reviews. 1976(8):25-47.
77. Sreebny LM. Saliva in health and disease: an appraisal and update. International dental journal. 2000;50(3):140-61.
78. Sreebny LM. Salivary flow in health and disease. Compendium (Newtown, Pa) Supplement. 1989(13):S461-9.
79. Turner RJ, Sugiya H. Understanding salivary fluid and protein secretion. Oral Diseases. 2002;8(1):3-11.
80. De Almeida PDV, Grégio AMT, Machado MÂN, De Lima AAS, Azevedo LR. Saliva composition and functions: A comprehensive review. J Contemp Dental Pract. 2008;9(3):072-80.
81. Humphrey SP, Williamson RT. A review of saliva: Normal composition, flow, and function. The Journal of Prosthetic Dentistry. 2001;85(2):162-9.
82. Grant DA, Stern IB, Listgarten MA. Saliva. Periodontics 6th Ed. 1988:135-46.
83. Dowd FJ. Saliva and dental caries. Dent Clin North Am. 1999;43(4):579-97.
84. Vitorino R, Barros A, Caseiro A, Domingues P, Duarte J, Amado F. Towards defining the whole salivary peptidome. Proteomics - Clinical Applications. 2009;3(5):528-40.
85. Vitorino R, Barros AS, Caseiro A, Ferreira R, Amado F. Evaluation of different extraction procedures for salivary peptide analysis. Talanta. 2012;94:209-15.
86. Hu S, Arellano M, Boontheung P, Wang J, Zhou H, Jiang J, et al. Salivary proteomics for oral cancer biomarker discovery. Clinical Cancer Research. 2008;14(19):6246-52.
87. de Jong EP, Xie H, Onsongo G, Stone MD, Chen X-B, Kooren JA, et al. Quantitative Proteomics Reveals Myosin and Actin as Promising Saliva Biomarkers for Distinguishing Pre-Malignant and Malignant Oral Lesions. PLoS ONE. 2010;5(6):e11148.
88. Xie H, Onsongo G, Popko J, de Jong EP, Cao J, Carlis JV, et al. Proteomics analysis of cells in whole saliva from oral cancer patients via value-added three-dimensional peptide fractionation and tandem mass spectrometry. Mol Cell Proteomics. 2008;7(3):486-98.
89. Griffin TJ, editor Human saliva proteome: an overview2014.
90. Cabras T, Pisano E, Boi R, Olianas A, Manconi B, Inzitari R, et al. Age-dependent modifications of the human salivary secretory protein complex. J Proteome Res. 2009;8(8):4126-34.
91. Castagnola M, Inzitari R, Fanali C, Iavarone F, Vitali A, Desiderio C, et al. The surprising composition of the salivary proteome of preterm human newborn. Mol Cell Proteomics. 2011;10(1):M110.003467.

92. Schipper RG, Silletti E, Vingerhoeds MH. Saliva as research material: Biochemical, physicochemical and practical aspects. Archives of Oral Biology. 2007;52(12):1114-35.
93. Tavassoli M, Brunel N, Maher R, Johnson NW, Soussi T. p53 antibodies in the saliva of patients with squamous cell carcinoma of the oral cavity [3]. International Journal of Cancer. 1998;78(3):390-1.
94. Xie H, Onsongo G, Popko J, de Jong EP, Cao J, Carlis JV, et al. Proteomics analysis of cells in whole saliva from oral cancer patients via value-added three-dimensional peptide fractionation and tandem mass spectrometry. Molecular and Cellular Proteomics. 2008;7(3):486-98.
95. Franzmann EJ, Reategui EP, Pedroso F, Pernas FG, Karakullukcu BM, Carraway KL, et al. Soluble CD44 is a potential marker for the early detection of head and neck cancer. Cancer Epidemiology Biomarkers and Prevention. 2007;16(7):1348-55.
96. Wu JY, Yi C, Chung HR, Wang DJ, Chang WC, Lee SY, et al. Potential biomarkers in saliva for oral squamous cell carcinoma. Oral Oncology. 2010;46(4):226-31.
97. Denny P, Hagen FK, Hardt M, Liao L, Yan W, Arellanno M, et al. The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. Journal of Proteome Research. 2008;7(5):1994-2006.
98. Li Y, St. John MAR, Zhou X, Kim Y, Sinha U, Jordan RCK, et al. Salivary transcriptome diagnostics for oral cancer detection. Clinical Cancer Research. 2004;10(24):8442-50.
99. Park NJ, Zhou H, Elashoff D, Henson BS, Kastratovic DA, Abemayor E, et al. Salivary microRNA: Discovery, characterization, and clinical utility for oral cancer detection. Clinical Cancer Research. 2009;15(17):5473-7.
100. Pow EHN, Law MYT, Tsang PCS, Perera RAPM, Kwong DLW. Salivary Epstein-Barr virus DNA level in patients with nasopharyngeal carcinoma following radiotherapy. Oral Oncology. 2011;47(9):879-82.
101. Xiao H, Wong DTW. Method development for proteome stabilization in human saliva. Analytica Chimica Acta. 2012;722:63-9.
102. Jiang J, Park NJ, Hu S, Wong DT. A universal pre-analytic solution for concurrent stabilization of salivary proteins, RNA and DNA at ambient temperature. Archives of Oral Biology. 2009;54(3):268-73.
103. Chaiyarit P, Utrawichian A, Leelayuwat C, Vatanasapt P, Chanchareonsook N, Samson MH, et al. Investigation of trefoil factor expression in saliva and oral mucosal tissues of patients with oral squamous cell carcinoma. Clinical Oral Investigations. 2012;16(6):1549-56.
104. Allegra E, Trapasso S, Sacco A, Aragona T, Belfiore A, Garozzo A. Elisa detection of salivary levels of cd44sol as a diagnostic test for laryngeal carcinomas. Journal of Cancer Science and Therapy. 2012;4(10):330-4.
105. Helmerhorst EJ, Oppenheim FG. Saliva: A dynamic proteome. Journal of Dental Research. 2007;86(8):680-93.
106. Messana I, Inzitari R, Fanali C, Cabras T, Castagnola M. Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us? Journal of Separation Science. 2008;31(11):1948-63.
107. Castagnola M, Cabras T, Vitali A, Sanna MT, Messana I. Biotechnological implications of the salivary proteome. Trends Biotechnol. 2011;29(8):409-18.

108. Chauncey HH, Lionetti F, Winer RA, Lisanti VF. Enzymes of Human Saliva: I. the Determination, Distribution, and Origin of whole Saliva Enzymes. Journal of Dental Research. 1954;33(3):321-34.
109. Trindade Fb, Amado F, Gomes PS, Vitorino R. endoProteoFASP: A novel FASP approach to profile salivary peptidome and disclose salivary proteases. Talanta. 2015;132(0):486-93.
110. Helmerhorst EJ. Whole Saliva Proteolysis. Annals of the New York Academy of Sciences. 2007;1098(1):454-60.
111. Fenton RG, Longo DL. Cancer cell biology and angiogenesis. Harrison's Principles of Internal Medicine. 2005:453-64.
112. Morin PJ, Trent JM, Collins FS, Vogelstein B. Cancer genetics. Harrison's Principles of Internal Medicine. 2005:447-53.
113. Evans AS, Mueller NE. Viruses and cancer. Causal associations. Annals of Epidemiology. 1990;1(1):71-92.
114. Psyrri A, DiMaio D. Human papillomavirus in cervical and head-and-neck cancer. Nature Clinical Practice Oncology. 2008;5(1):24-31.
115. Steben M, Duarte-Franco E. Human papillomavirus infection: Epidemiology and pathophysiology. Gynecologic Oncology. 2007;107(2 SUPPL.):S2-S5.
116. Johnson NW. Risk Markers for Oral Disease, Oral Cancer Detection of Patients and Lesions at Risk. 1991;2.
117. Nagao Y, Sata M, Tanikawa K, Itoh K, Kameyama T. High prevalence of hepatitis C virus antibody and RNA in patients with oral cancer. Journal of Oral Pathology and Medicine. 1995;24(8):354-60.
118. Porter SR, Lodi G, Chandler K, Kumar N. Development of squamous cell carcinoma in hepatitis C virus-associated lichen planus. European Journal of Cancer Part B: Oral Oncology. 1997;33(1):58-9.
119. Colson P, La Scola B, Raoult D. Giant Viruses of Amoebae as Potential Human Pathogens. Intervirology. 2013;56(6):376-85.
120. Bernard La S, Thomas JM, Jean-Pierre A, Didier R. Mimivirus in Pneumonia Patients. Emerging Infectious Disease journal. 2005;11(3):449.
121. Pierre B, Laurent P, Michel D, Bernard La S, Jean-Pierre A, Didier R. Ameba-associated Microorganisms and Diagnosis of Nosocomial Pneumonia. Emerging Infectious Disease journal. 2006;12(2):248.
122. Hu J, Dong CY, Li JKK, Chen DE, Liang K, Liu J. Selective in vitro cytotoxic effect of human cancer cells by Bluetongue virus-10. Acta Oncologica. 2008;47(1):124-34.
123. Yang HY, Dong CY, Liu J, Li XC, Deng GL, Bi GM. Comparative analysis of targeted oncolysis of five strain bluetongue virus. Chinese Journal of Microbiology and Immunology. 2010;30(7):621-5.
124. Li JK. Oncolytic Bluetongue Viruses: Promise, Progress and Perspectives. Frontiers in Microbiology. 2011;2.
125. Marvil P, Knowles NJ, Mockett APA, Britton P, Brown TDK, Cavanagh D. Avian encephalomyelitis virus is a picornavirus and is most closely related to hepatitis A virus. Journal of General Virology. 1999;80(3):653-62.
126. Hughes PJ, Stanway G. The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation. Journal of General Virology. 2000;81(1):201-7.

127. Boya P, Pauleau AL, Poncet D, Gonzalez-Polo RA, Zamzami N, Kroemer G. Viral proteins targeting mitochondria: Controlling cell death. Biochimica et Biophysica Acta - Bioenergetics. 2004;1659(2-3):178-89.
128. Witter RL. Avian tumor viruses: Persistent and evolving pathogens. Acta Veterinaria Hungarica. 1997;45(3):251-66.
129. Kumar S, Kunec D, Buza JJ, Chiang HI, Zhou H, Subramaniam S, et al. Nuclear Factor kappa B is central to Marek's Disease herpesvirus induced neoplastic transformation of CD30 expressing lymphocytes in-vivo. BMC Systems Biology. 2012;6.
130. Samorek-Salamonowicz E, Czekaj H, Kozdrun W. Marek's disease herpes virus - A potential risk for human health. Medycyna Weterynaryjna. 2003;59(4):293-6.
131. Deng X, Li X, Shen Y, Qiu Y, Shi Z, Shao D, et al. The Meq oncoprotein of Marek's disease virus interacts with p53 and inhibits its transcriptional and apoptotic activities. Virology Journal. 2010;7.
132. Hooker CW, Scott J, Apolloni A, Parry E, Harrich D. Human immunodeficiency virus type 1 reverse transcription is stimulated by Tat from other lentiviruses. Virology. 2002;300(2):226-35.
133. Bhatia S, Patil SS, Sood R. Bovine immunodeficiency virus: A lentiviral infection. Indian Journal of Virology. 2013;24(3):332-41.
134. Brun A, Rivas C, Esteban M, Escribano JM, Alonso C. African swine fever virus gene A179L, a viral homologue of bcl-2, protects cells from programmed cell death. Virology. 1996;225(1):227-30.
135. Revilla Y, Cebrián A, Baixerás E, Martínez-A C, Viñuela E, Salas ML. Inhibition of apoptosis by the African swine fever virus Bcl-2 homologue: Role of the BH1 domain. Virology. 1997;228(2):400-4.
136. Cholleti H, Paidikondala M, Munir M, Hakhverdyan M, Baule C. Equine arteritis virus induced cell death is associated with activation of the intrinsic apoptotic signalling pathway. Virus Research. 2013;171(1):222-6.
137. St-Louis MC, Archambault D. The equine arteritis virus induces apoptosis via caspase-8 and mitochondria-dependent caspase-9 activation. Virology. 2007;367(1):147-55.
138. Zhang J, Timoney PJ, MacLachlan NJ, McCollum WH, Balasuriya UBR. Persistent equine arteritis virus infection in HeLa cells. Journal of Virology. 2008;82(17):8456-64.
139. Caughman GB, Staczek J, O'Callaghan DJ. Equine cytomegalovirus: Structural proteins of virions and nucleocapsids. Virology. 1984;134(1):184-95.
140. Staczek J, Wharton JH, Dauenhauer SA, O'Callaghan DJ. Coestablishment of persistent infection and oncogenic transformation of hamster embryo cells by equine cytomegalovirus. Virology. 1984;132(2):339-51.
141. Sharma PC, Cullinane AA, Onions DE, Nicolson L. Diagnosis of equid herpesviruses -1 and -4 by polymerase chain reaction. Equine veterinary journal. 1992;24(1):20-5.
142. Ramqvist T, Dalianis T. Lessons from immune responses and vaccines against Murine polyomavirus infection and polyomavirus-induced tumours potentially useful for studies on human polyomaviruses. Anticancer Research. 2010;30(2):279-84.
143. Nicol JTJ, Touzé A. Prevalence of polyomaviruses and links with human diseases. Future Virology. 2011;6(10):1187-97.

144. Ikeda Y, Tomonaga K, Kawaguchi Y, Kohmoto M, Inoshima Y, Tohya Y, et al. Feline immunodeficiency virus can infect a human cell line (MOLT-4) but establishes a state of latency in the cells. Journal of General Virology. 1996;77(8):1623-30.
145. Vannucci L, Chiuppesi F, di Martino F, Caligo MA, Bevilacqua G, Pistello M. Feline immunodeficiency virus vector as a tool for preventative strategies against human breast cancer. Veterinary Immunology and Immunopathology. 2010;134(1-2):132-7.
146. Bester R, Pepler PT, Burger JT, Maree HJ. Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. Journal of Virological Methods. 2014;210:67-75.
147. Cogotzi L, Giampetruzzi A, Nölke G, Orecchia M, Elicio V, Castellano MA, et al. An assay for the detection of grapevine leafroll-associated virus 3 using a single-chain fragment variable antibody. Archives of Virology. 2009;154(1):19-26.
148. Matthews REF, Ralph RK. Turnip Yellow Mosaic Virus. Advances in Virus Research1967. p. 273-328.
149. Shah R, Coleman CS, Mir K, Baldwin J, Van Etten JL, Grishin NV, et al. Paramecium bursaria Chlorella Virus-1 Encodes an Unusual Arginine Decarboxylase That Is a Close Homolog of Eukaryotic Ornithine Decarboxylases. Journal of Biological Chemistry. 2004;279(34):35760-7.
150. Yamada T, Onimatsu H, Van Etten JL. Chlorella Viruses. Advances in Virus Research2006. p. 293-336.
151. Xu LJ, Jiang T, Zhang FJ, Han JF, Liu J, Zhao H, et al. Global Transcriptomic Analysis of Human Neuroblastoma Cells in Response to Enterovirus Type 71 Infection. PLoS ONE. 2013;8(7).
152. Tapisiz A, Polat M, Kara SS, Tezer H, Simsek H, Aktas F. Prevention of measles spread on a paediatric ward. Epidemiology and Infection. 2014.
153. Raekiansyah M, Espada-Murao LA, Okamoto K, Kubo T, Morita K. Dengue virus neither directly mediates hyperpermeabi lity nor enhances tumor necrosis factor-α-induced permeability in vitro. Japanese Journal of Infectious Diseases. 2014;67(2):86-94.
154. Margine I, Krammer F. Animal models for influenza viruses: Implications for universal vaccine development. Pathogens. 2014;3(4):846-74.
155. Husain M. Avian influenza A (H7N9) virus infection in humans: Epidemiology, evolution, and pathogenesis. Infection, Genetics and Evolution. 2014;28:304-12.
156. Khachaturov V, Xiao G-Q, Kinoshita Y, Unger PD, Burstein DE. Histone H1.5, a novel prostatic cancer marker: an immunohistochemical study. Human Pathology. 2014;45(10):2115-9.
157. Lu A, Zougman A, Pudełko M, Bębenek M, Ziółkowski P, Mann M, et al. Mapping of Lysine Monomethylation of Linker Histones in Human Breast and Its Cancer. Journal of Proteome Research. 2009;8(9):4207-15.
158. Olsson JCA-K. Platelet Regulation of Angiogenesis, Tumor Growth and Metastasis, Tumor Angiogenesis. 2012.
159. Vairaktaris E, Serefoglou Z, Yapijakis C, Stathopoulos P, Vassiliou S, Derka S, et al. The interleukin-1 beta gene polymorphism +3953 C/T is not associated with risk for oral cancer. Anticancer Research. 2007;27(6 B):3981-6.
160. Kamatani T, Shiogama S, Yoshihama Y, Kondo S, Shirota T, Shintani S. Interleukin-1 beta in unstimulated whole saliva is a potential biomarker for oral squamous cell carcinoma. Cytokine. 2013;64(2):497-502.

161. Almeida LO, Abrahao AC, Rosselli-Murai LK, Giudice FS, Zagni C, Leopoldino AM, et al. NFκB mediates cisplatin resistance through histone modifications in head and neck squamous cell carcinoma (HNSCC). FEBS Open Bio. 2014;4(0):96-104.
162. Chen YW, Kao SY, Wang HJ, Yang MH. Histone modification patterns correlate with patient outcome in oral squamous cell carcinoma. Cancer. 2013;119(24):4259-67.
163. Monteiro FL, Baptista T, Amado F, Vitorino R, Jerónimo C, Helguero LA. Expression and functionality of histone H2A variants in cancer. Oncotarget. 2014;5(11):3428-43.
164. Schmidt M, Polednik C, Hoppe F. Proteolytic patterns of head and neck squamous cell carcinoma. European Archives of Oto-Rhino-Laryngology. 1999;256(7):346-50.

2. ATTACHMENTS:

2.1 Description of the proteins found and respective acronym.

Protein	Description
6PGD_HUMAN	6-phosphogluconate dehydrogenase, decarboxylation
ACTA_HUMAN	Actin, aortic smooth muscle
ACTB_HUMAN	Actin, cytoplasmic
ACTG_HUMAN	Actin, cytoplasmic 2
ALBU_HUMAN	Serum albumin
AMY1_HUMAN	Alpha-amylase 1
AMYP_HUMAN	Pancreatic alpha-amylase
ANXA1_HUMAN	Annexin A1
APOA1_HUMAN	Apolipoprotein A-I
ATOH1_HUMAN	Protein atonal homolog 1
B2MG_HUMAN	Beta-2-microglobulin
BPIA2_HUMAN	BPI fold-containing family A member 2
BPI_HUMAN	Bactericidal permeability-increasing protein
BPIB1_HUMAN	BPI fold-containing family B member 1
BRPF3_HUMAN	Bromodomain and PHD finger-containing protein 3
CAH6_HUMAN	Carbonic anhydrase 6
CAP7_HUMAN	Azurocidin
CENPV_HUMAN	Centromere protein V
CATG_HUMAN	Cathepsin G
CD040_HUMAN	Uncharacterized protein
CO1A1_HUMAN	Collagen alpha-1 (I) chain
COBL_HUMAN	Protein cordon-bleu
CRNN_HUMAN	Cornulin
CYTB_HUMAN	Cystatin-B
CYTN_HUMAN	Cystatin-SN
ELNE_HUMAN	Neutrophil elastase
ENOA_HUMAN	Alpha-enolase
EZRI_HUMAN	Ezrin
FIBA_HUMAN	Fibrinogen alpha chain
FIBB_HUMAN	Fibrinogen beta chain
G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase
H11_HUMAN	Histone H1,1
H12_HUMAN	Histone H1,2

H13_HUMAN	Histone H1,3
H14_HUMAN	Histone H1,4
H15_HUMAN	Histone H1,5
H2A1B_HUMAN	Histone H2A type 1-B/E
H2A1C_HUMAN	Histone H2A type 1-C
H2A1H_HUMAN	Histone H2A type 1-H
H2A2A_HUMAN	Histone H2A type 2-A
H2A2B_HUMAN	Histone H2A type 2-B
H2A2C_HUMAN	Histone H2A type 2-C
H2AJ_HUMAN	Histone H2A,J
H2B1B_HUMAN	Histone H2B type 1-B
H2B1C_HUMAN	Histone H2B type 1-C/E/F/G/I
H2B1D_HUMAN	Histone H2B type 1-D
H2B1H_HUMAN	Histone H2B type 1-H
H2B1J_HUMAN	Histone H2B type 1-J
H2B1K_HUMAN	Histone H2B type 1-K
H2B1M_HUMAN	Histone H2B type 1-M
H2B1O_HUMAN	Histone H2B type 1-O
H2B2E_HUMAN	Histone H2B type 2-E
H2B3B_HUMAN	Histone H2B type 3-B
H2BFS_HUMAN	Histone H2B type F-S
H31_HUMAN	Histone H3,1
H31T_HUMAN	Histone H3,1t
H4_HUMAN	Histone H4
HBA_HUMAN	Hemoglobin subunit alpha
HBB_HUMAN	Hemoglobin subunit beta
HIS1_HUMAN	Histatin-1
HIS3_HUMAN	Histatin-3
HMGB2_HUMAN	High mobility group protein B2
HMGN2_HUMAN	Non-histone chromosomal protein HMG-17
HNRL1_HUMAN	Heterogeneous nuclear ribonucleoprotein U-like protein 1
HS902_HUMAN	Putative heat shock protein HSP 90-alpha A2
HSPB1_HUMAN	Heat shock protein beta-1
INF2_HUMAN	Inverted formin-2
K1C13_HUMAN	Keratin, type I cytoskeletal 13
K2C4_HUMAN	Keratin, type II cytoskeletal 4
K2C6A_HUMAN	Keratin, type II cytoskeletal 6A

K2C6B_HUMAN	Keratin, type II cytoskeletal 6B
LEG7_HUMAN	Galectin-7
MAMD1_HUMAN	Mastermind-like domain-containing protein 1
MNDA_HUMAN	Myeloid cell nuclear differentiation antigen
MUC7_HUMAN	Mucin-7
NAV2_HUMAN	Neuron navigator 2
PIGR_HUMAN	Polymeric immunoglobulin receptor
PIP_HUMAN	Prolactin-inducible protein
PRB2_HUMAN	Basic salivary proline-rich protein 2
PRB3_HUMAN	Basic salivary proline-rich protein 3
PRB4_HUMAN	Basic salivary proline-rich protein 4
PROL4_HUMAN	Proline-rich protein 4
PRP1_HUMAN	Basic salivary proline-rich protein 1
PRPC_HUMAN	Salivary acidic proline-rich phosphoprotein 1/2
S10A8_HUMAN	Protein S100-A8
S10A9_HUMAN	Protein S100-A9
S10AC_HUMAN	Protein S100-A12
SET1B_HUMAN	Histone-lysine N-methyltransferase SETD1B
SMCA5_HUMAN	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5
SMR3B_HUMAN	Submaxillary gland androgen-regulated protein 3B
STAT_HUMAN	Statherin
THIO_HUMAN	Thioredoxin
TRFL_HUMAN	Lactotransferrin
TYB4_HUMAN	Thymosin beta-4
VASP_HUMAN	Vasodilator-stimulated phosphoprotein
ZG16B_HUMAN	Zymogen granule protein 16 homolog B
SPRR3_HUMAN	Small proline-rich protein 3

Table of the identified proteins in at least 2 individuals by group

Proteins in Control Group (30)	Proteins in Larynx Group (66)	Proteins in Pharynx and Oral Cavity Group (53)	Proteins in Nasopharynx and Nostrils Group (24)
ACTB_HUMAN	ACTB_HUMAN	ACTB_HUMAN	ACTB_HUMAN
ALBU_HUMAN	ACTG_HUMAN	ACTBL_HUMAN	CATG_HUMAN
AMY1_HUMAN	AMY1_HUMAN	ALDOA_HUMAN	FIBB_HUMAN
ANXA1_HUMAN	AMYP_HUMAN	AMY1_HUMAN	G3P_HUMAN
BPIB1_HUMAN	ANXA1_HUMAN	APOA1_HUMAN	H15_HUMAN
CATG_HUMAN	ATOH1_HUMAN	BPIB1_HUMAN	H2A2A_HUMAN
CRNN_HUMAN	B2MG_HUMAN	BRPF3_HUMAN	H4_HUMAN
CYTB_HUMAN	BPIA2_HUMAN	CAP1_HUMAN	HBA_HUMAN
CYTN_HUMAN	BPIB1_HUMAN	CATG_HUMAN	HBB_HUMAN
FIBB_HUMAN	CAH6_HUMAN	CD040_HUMAN	HIS1_HUMAN
G3P_HUMAN	CAP7_HUMAN	COBL_HUMAN	MUC7_HUMAN
H2A2A_HUMAN	CENPV_HUMAN	CRNN_HUMAN	PIGR_HUMAN
H2B1D_HUMAN	CATG_HUMAN	FIBA_HUMAN	PRB2_HUMAN
H4_HUMAN	CD040_HUMAN	FIBB_HUMAN	PRB3_HUMAN
HBB_HUMAN	CO1A1_HUMAN	G3P_HUMAN	PRB4_HUMAN
HIS1_HUMAN	CRNN_HUMAN	H12_HUMAN	PRP1_HUMAN
K1C13_HUMAN	CYTN_HUMAN	H13_HUMAN	PRPC_HUMAN
MUC7_HUMAN	ELNE_HUMAN	H15_HUMAN	S10A9_HUMAN
PIP_HUMAN	EZRI_HUMAN	H2A1C_HUMAN	SMR3B_HUMAN
PRB2_HUMAN	FIBB_HUMAN	H2A2A_HUMAN	STAT_HUMAN
PRB3_HUMAN	G3P_HUMAN	H2AJ_HUMAN	TRFL_HUMAN
PRB4_HUMAN	H12_HUMAN	H2B1B_HUMAN	TYB4_HUMAN
PRP1_HUMAN	H14_HUMAN	H2B1D_HUMAN	ZG16B_HUMAN
PRPC_HUMAN	H15_HUMAN	H2B1H_HUMAN	SPRR3_HUMAN
S10A9_HUMAN	H2A1C_HUMAN	H2B1O_HUMAN	
SMR3B_HUMAN	H2A1H_HUMAN	H31_HUMAN	
STAT_HUMAN	H2A2A_HUMAN	H4_HUMAN	
THIO_HUMAN	H2B1B_HUMAN	HBA_HUMAN	
TRFL_HUMAN	H2B1M_HUMAN	HBB_HUMAN	
ZG16B_HUMAN	H2B2E_HUMAN	HBD_HUMAN	
	H2BFS_HUMAN	HIS1_HUMAN	
	H31_HUMAN	HIS3_HUMAN	
	H31T_HUMAN	HMGB2_HUMAN	
	H4_HUMAN	LEG7_HUMAN	

HBA_HUMAN	MNDA_HUMAN
HBB_HUMAN	MUC7_HUMAN
HIS1_HUMAN	PRB2_HUMAN
HIS3_HUMAN	PRB3_HUMAN
HMGN2_HUMAN	PRB4_HUMAN
HNRL1_HUMAN	PROL4_HUMAN
HSPB1_HUMAN	PRP1_HUMAN
K2C4_HUMAN	PRPC_HUMAN
K2C6B_HUMAN	PRC2C_HUMAN
LEG7_HUMAN	S10A8_HUMAN
MAMD1_HUMAN	S10A9_HUMAN
MNDA_HUMAN	S10AC_HUMAN
MUC7_HUMAN	S2541_HUMAN
PIGR_HUMAN	SET1B_HUMAN
PIP_HUMAN	SMR3B_HUMAN
PRB2_HUMAN	STAT_HUMAN
PRB3_HUMAN	TRFL_HUMAN
PRB4_HUMAN	TYB4_HUMAN
PROL4_HUMAN	ZG16B_HUMAN
PRP1_HUMAN	
PRPC_HUMAN	
S10A8_HUMAN	
S10A9_HUMAN	
S10AC_HUMAN	
SMCA5_HUMAN	
SMR3B_HUMAN	
STAT_HUMAN	
TRFL_HUMAN	
TYB4_HUMAN	
VASP_HUMAN	
ZG16B_HUMAN	
SPRR3_HUMAN	

Venn Diagrams showing the distribution of the identified proteins between the Control group and the Larynx group

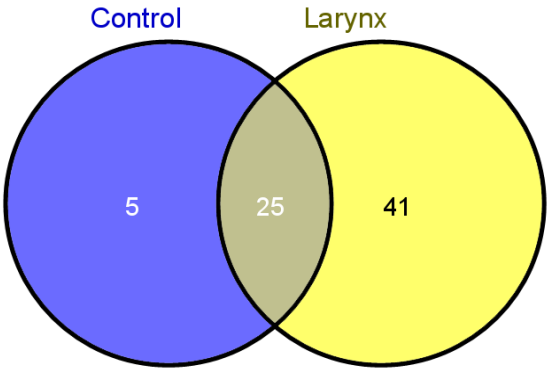


Table of the distribution of the identified proteins between the Control group and the Larynx group

Proteins only in Control Group (5)	Proteins in Control Group and Larynx Group (25)	Proteins only in Larynx Group (41)
LBU_HUMAN	ACTB_HUMAN	ACTG_HUMAN
CYTB_HUMAN	AMY1_HUMAN	AMYP_HUMAN
H2B1D_HUMAN	ANXA1_HUMAN	ATOH1_HUMAN
K1C13_HUMAN	BPIB1_HUMAN	B2MG_HUMAN
THIO_HUMAN	CATG_HUMAN	BPIA2_HUMAN
	CRNN_HUMAN	CAH6_HUMAN
	CYTN_HUMAN	CAP7_HUMAN
	FIBB_HUMAN	CENPV_HUMAN
	G3P_HUMAN	CD040_HUMAN
	H2A2A_HUMAN	CO1A1_HUMAN
	H4_HUMAN	ELNE_HUMAN
	HBB_HUMAN	EZRI_HUMAN
	HIS1_HUMAN	H12_HUMAN
	MUC7_HUMAN	H14_HUMAN
	PIP_HUMAN	H15_HUMAN
	PRB2_HUMAN	H2A1C_HUMAN
	PRB3_HUMAN	H2A1H_HUMAN
	PRB4_HUMAN	H2B1B_HUMAN

	PRP1_HUMAN	H2B1M_HUMAN
	PRPC_HUMAN	H2B2E_HUMAN
	S10A9_HUMAN	H2BFS_HUMAN
	SMR3B_HUMAN	H31_HUMAN
	STAT_HUMAN	H31T_HUMAN
	TRFL_HUMAN	HBA_HUMAN
	ZG16B_HUMAN	HIS3_HUMAN
		HMGN2_HUMAN

Venn Diagrams showing the distribution of the identified proteins between the Control group and the Pharynx + Oral Cavity group

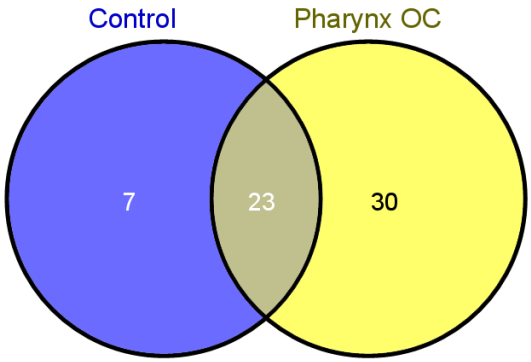


Table of the distribution of the identified proteins between the Control group and the Pharynx and Oral Cavity group

Proteins only in Control Group (7)	Proteins in Control Group and Pharynx and Oral Cavity Group (23)	Proteins only in Pharynx + Oral Cavity (30)
ALBU_HUMAN	ACTB_HUMAN	ACTBL_HUMAN
ANXA1_HUMAN	AMY1_HUMAN	ALDOA_HUMAN
CYTB_HUMAN	BPIB1_HUMAN	APOA1_HUMAN
CYTN_HUMAN	CATG_HUMAN	BRPF3_HUMAN
K1C13_HUMAN	CRNN_HUMAN	CAP1_HUMAN
PIP_HUMAN	FIBB_HUMAN	CD040_HUMAN
THIO_HUMAN	G3P_HUMAN	COBL_HUMAN
	H2A2A_HUMAN	FIBA_HUMAN
	H2B1D_HUMAN	H12_HUMAN
	H4_HUMAN	H13_HUMAN
	HBB_HUMAN	H15_HUMAN
	HIS1_HUMAN	H2A1C_HUMAN
	MUC7_HUMAN	H2AJ_HUMAN
	PRB2_HUMAN	H2B1B_HUMAN
	PRB3_HUMAN	H2B1H_HUMAN
	PRB4_HUMAN	H2B1O_HUMAN
	PRP1_HUMAN	H31_HUMAN

	PRPC_HUMAN	HBA_HUMAN
	S10A9_HUMAN	HBD_HUMAN
	SMR3B_HUMAN	HIS3_HUMAN
	STAT_HUMAN	HMGB2_HUMAN
	TRFL_HUMAN	LEG7_HUMAN
	ZG16B_HUMAN	MNDA_HUMAN
		PROL4_HUMAN
		PRC2C_HUMAN
		S10A8_HUMAN
		S10AC_HUMAN
		S2541_HUMAN
		SET1B_HUMAN
		TYB4_HUMAN

Venn Diagrams showing the distribution of the identified proteins between the Control group and the Nasopharynx and Nostrils Group

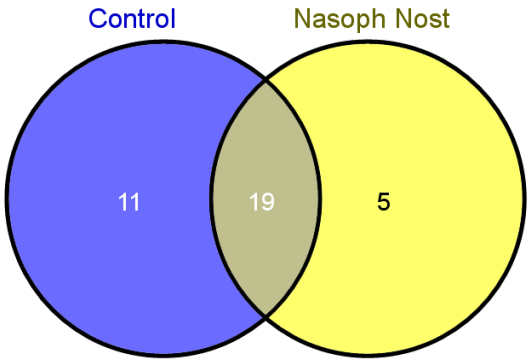
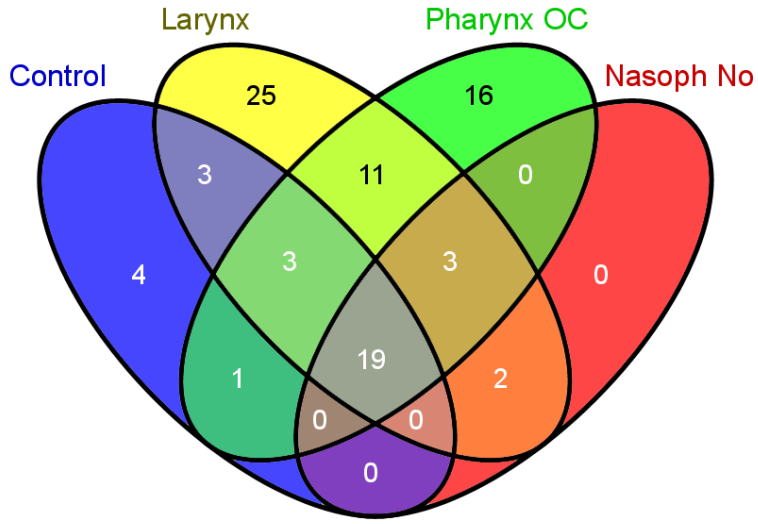


Table of the distribution of the identified proteins between the Control group and the Pharynx and Nasopharynx and Nostrils group

Proteins only in Control Group (11)	Proteins in Control Group and Nasopharynx and Nostrils Group (19)	Proteins only in Nasopharynx and Nostrils Group (5)
ALBU_HUMAN	ACTB_HUMAN	H15_HUMAN
AMY1_HUMAN	CATG_HUMAN	HBA_HUMAN
ANXA1_HUMAN	FIBB_HUMAN	PIGR_HUMAN
BPIB1_HUMAN	G3P_HUMAN	TYB4_HUMAN
CRNN_HUMAN	H2A2A_HUMAN	SPRR3_HUMAN
CYTB_HUMAN	H4_HUMAN	
CYTN_HUMAN	HBB_HUMAN	
H2B1D_HUMAN	HIS1_HUMAN	
K1C13_HUMAN	MUC7_HUMAN	
PIP_HUMAN	PRB2_HUMAN	
THIO_HUMAN	PRB3_HUMAN	
	PRB4_HUMAN	
	PRP1_HUMAN	
	PRPC_HUMAN	
	S10A9_HUMAN	
	SMR3B_HUMAN	
	STAT_HUMAN	
	TRFL_HUMAN	
	ZG16B_HUMAN	

2.2 Venn Diagrams showing the distribution of the identified proteins between the Control group, the Larynx group, the Pharynx and Oral Cavity group and the Nasopharynx and Nostrils group

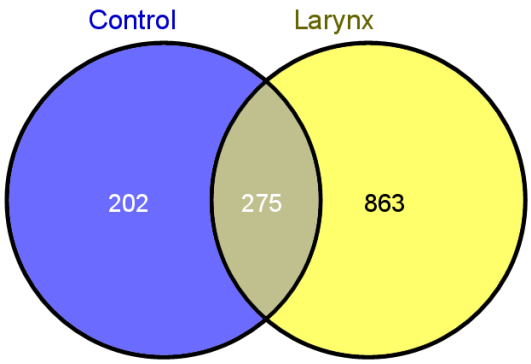


2.3 Table of the distribution of the identified proteins between Control group, the Larynx group, the Pharynx and Oral Cavity group and the Nasopharynx and Nostrils group

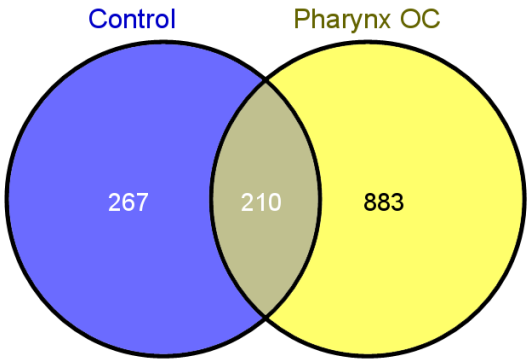
Proteins only in Control Group (4)	Proteins only in Larynx Group (25)	Proteins only in Pharynx and Oral Cavity Group (16)	Proteins common to all groups (19)	Proteins common to all cancer groups (3)
ALBU_HUMAN	ACTG_HUMAN	ACTBL_HUMAN	ACTB_HUMAN	H15_HUMAN
CYTB_HUMAN	AMYP_HUMAN	ALDOA_HUMAN	CATG_HUMAN	HBA_HUMAN
K1C13_HUMAN	ATOH1_HUMAN	APOA1_HUMAN	FIBB_HUMAN	TYB4_HUMAN
THIO_HUMAN	B2MG_HUMAN	BRPF3_HUMAN	G3P_HUMAN	
	BPIA2_HUMAN	CAP1_HUMAN	H2A2A_HUMAN	
	CAH6_HUMAN	COBL_HUMAN	H4_HUMAN	
	CAP7_HUMAN	FIBA_HUMAN	HBB_HUMAN	
	CENPV_HUMAN	H13_HUMAN	HIS1_HUMAN	
	CO1A1_HUMAN	H2AJ_HUMAN	MUC7_HUMAN	
	ELNE_HUMAN	H2B1H_HUMAN	PRB2_HUMAN	
	EZRI_HUMAN	H2B1O_HUMAN	PRB3_HUMAN	
	H14_HUMAN	HBD_HUMAN	PRB4_HUMAN	
	H2A1H_HUMAN	HMGB2_HUMAN	PRP1_HUMAN	

H2B1M_HUMAN	PRC2C_HUMAN	PRPC_HUMAN
H2B2E_HUMAN	S2541_HUMAN	S10A9_HUMAN
H2BFS_HUMAN	SET1B_HUMAN	SMR3B_HUMAN
H31T_HUMAN		STAT_HUMAN
HMGN2_HUMAN		TRFL_HUMAN
HNRL1_HUMAN		ZG16B_HUMAN
HSPB1_HUMAN		
K2C4_HUMAN		
K2C6B_HUMAN		
MAMD1_HUMAN		
SMCA5_HUMAN		
VASP_HUMAN		

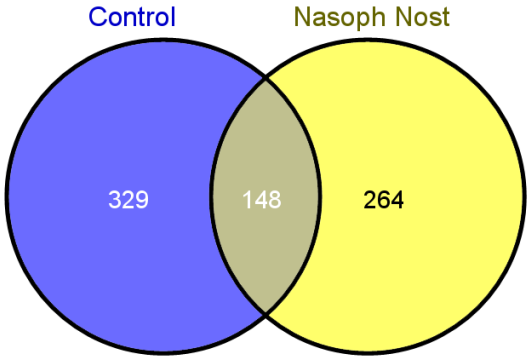
Venn Diagrams showing the distribution of the identified peptides between the Control group and the Larynx group



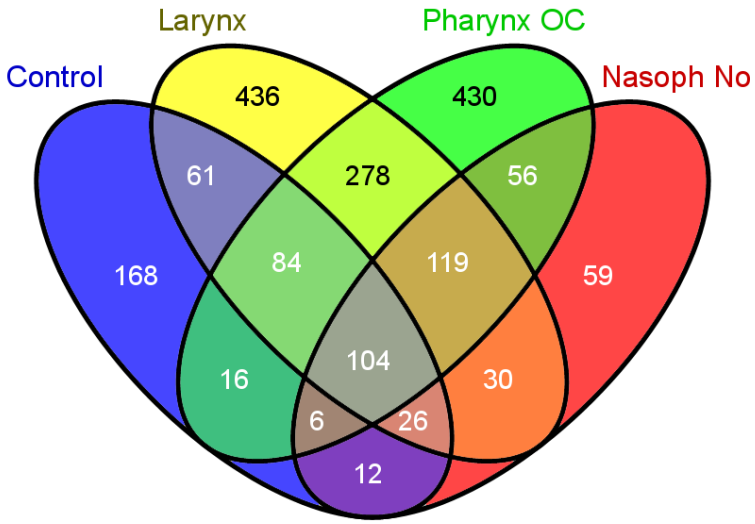
Venn Diagrams showing the distribution of the identified peptides between the Control group and the Pharynx and Oral Cavity group



Venn Diagrams showing the distribution of the identified peptides between the Control group and the Nasopharynx and Nostrils group



2.4 Venn Diagrams showing the distribution of the identified peptides between Control group, the Larynx group, the Pharynx and Oral Cavity group and the Nasopharynx and Nostrils group



2.5 Proteases from the control terminal residues

Terminal N Residue	Clivages Ratio	Peptidases
Asp Gly	0,0281	pepsin A; Cathepsin B, Caspase 1, HIV-1 retropepsin
Gln Gly	0,0276	Cathepsin L,S,K
Tyr Gly	0,0244	Pepsina A; Cathepsin D,L,S,B
Val Arg	0,0218	DegP peptidase; Cathepsin L; chymosin;rhizopuspepsin
Ile Ala	0,0207	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Met Asn	0,0164	granzyme M, meprin beta subunit, Cathepsin B, E, L, K, S
Gly Pro	0,0122	aminopeptidase P, cathepsin B, L3, matrix metallopeptidase, ubiquitin-specific peptidase 4 and 15
Ile Arg	0,0111	pepsin A, rhizopuspepsin, cathepsin L, calpain-1 and 2, elastase-2
Pro Pro	0,0111	Cathepsin L1 and L3, lactocepin I, calpain 1 and 2
Ser Asp	0,0095	pepsin A, aspergillopepsin I, necepsin 1 and 2, cathepsin L and S
Phe Gly	0,0090	pepsin A and B, cathepsin D, E and L, necepsin 1, HIV-1 retropepsin, chymotrypsin A

Terminal N Residue	Clivages Ratio	Peptidases
Asp Gly	0,0281	pepsin A; Cathepsin B, Caspase 1, HIV-1 retropepsin
Gln Gly	0,0276	Cathepsin L,S,K
Tyr Gly	0,0244	Pepsina A; Cathepsin D,L,S,B
Val Arg	0,0218	DegP peptidase; Cathepsin L; chymosin;rhizopuspepsin
Ile Ala	0,0207	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Met Asn	0,0164	granzyme M, meprin beta subunit, Cathepsin B, E, L, K, S
Gly Pro	0,0122	aminopeptidase P, cathepsin B, L3, matrix metallopeptidase, ubiquitin-specific peptidase 4 and 15
Ile Arg	0,0111	pepsin A, rhizopuspepsin, cathepsin L, calpain-1 and 2, elastase-2
Pro Pro	0,0111	Cathepsin L1 and L3, lactocepin I, calpain 1 and 2
Ser Asp	0,0095	pepsin A, aspergillopepsin I, necepsin 1 and 2, cathepsin L and S
Phe Gly	0,0090	pepsin A and B, cathepsin D, E and L, necepsin 1, HIV-1 retropepsin, chymotrypsin A

2.6 Proteases from the terminal residues by cancer

Laryngeal SCC

Terminal N Residue	Clivages Ratio	Peptidases
Gln Gly	0,1427	Cathepsin L,S,K
Met Asn	0,0621	granzyme M, meprin beta subunit, Cathepsin B, E, L, K, S
Tyr Gly	0,0605	Pepsina A; Cathepsin D,L,S,B
Arg Gly	0,0361	rhizopuspepsin, memapsin-1, yapsin A, cathepsin B, L and K
Phe Gly	0,0334	pepsin A and B, cathepsin D, E and L, necepsin 1, HIV-1 retropepsin, chymotrypsin A
Ile Met	0,0308	elastase-2, matrix metallopeptidase-2
Asp Gly	0,0297	pepsin A; Cathepsin B, Caspase 1, HIV-1 retropepsin
Pro Pro	0,0281	Cathepsin L1 and L3, lactocepin I, calpain 1 and 2
Lys Ser	0,0255	pepsin A, cathepsin D, E and L, necepsin-1, calpain 1 and 2
Ser Gln	0,0249	calpain 1 and 2, Cathepsin L and , matrix metallopeptidase-2

Terminal C Residue	Clivages Ratio	Peptidases
Gln Gly	0,1655	Cathepsin L,S,K
Tyr Gly	0,0796	Pepsina A; Cathepsin D,L,S,B
Ile Ala	0,0573	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Leu Ala	0,0419	Pepsin A, Cathepsin D and E
Lys Ala	0,0393	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Pro Arg	0,0361	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Gln Ser	0,0350	pepsin A, gastricsin, Cathepsin B, L and S
Ala Gly	0,0324	cathepsin L, S, K and B
Gln Tyr	0,0292	Cathepsin E, L, S and B, matrix metallopeptidase-2

Vocal Chords SCC

Terminal N Residue	Clivages Ratio	Peptidases
Tyr Val	0,0175	pepsin A, rhizopuspepsin, cathepsin D and E, HIV-1 retropepsin
Gln Gly	0,0159	Cathepsin L,S,K
Lys Ala	0,0133	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Ile Ala	0,0122	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Tyr Gly	0,0117	Pepsina A; Cathepsin D,L,S,B
Ile Met	0,0117	elastase-2, matrix metallopeptidase-2
Met Val	0,0080	cathepsin B, D, E, L, S, rhizopuspepsin
Leu Arg	0,0069	necepsin-1, presenilin 1, cathepsin D, E, L, H, matrix metallopeptidase-2,carboxypeptidase E,M, N, U
Phe Gly	0,0064	pepsin A and B, cathepsin D, E and L, necepsin 1, HIV-1 retropepsin, chymotrypsin A
Leu Ala	0,0058	Pepsin A, Cathepsin D and E

Terminal C Residue	Clivages Ratio	Peptidases
Ile Ala	0,0308	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Lys Ala	0,0233	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Tyr Gly	0,0133	Pepsina A; Cathepsin D,L,S,B
Ala His	0,0122	necepsin-2, Cathepsin L and S, falcipain 2 and 3
Ile Thr	0,0117	pepsin A, Cathepsin E, L and G, rhizopuspepsin, elastase-2, ERAP2 aminopeptidase
Val Tyr	0,0111	falcipain2 and 3, matrix metallopeptidase 2 and 13, HtrA2 peptidase, elastase-2
Gln Gly	0,0095	Cathepsin L,S,K
Pro Arg	0,0074	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Val Val	0,0069	thermolysin, leishmanolysin, elastase-2, lactocepin I and 3, cathepsin E, L1,

Mouth Floor SCC

Terminal N Residue	Clivages Ratio	Peptidases
Tyr Val	0,0133	pepsin A, rhizopuspepsin, cathepsin D and E, HIV-1 retropepsin
Gln Gly	0,0111	Cathepsin L,S,K
Met Val	0,0095	cathepsin B, D, E, L, S, rhizopuspepsin
Lys Ala	0,0080	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Ile Ala	0,0074	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Arg Ser	0,0074	yapsin-1 and 3, cathepsin B, L, S, K, P, rhodesain, gingipain R
Met Gly	0,0074	aspergillopepsin I, cathepsin B, L, S, K,
Tyr Gly	0,0069	Pepsina A; Cathepsin D,L,S,B
Ala Met	0,0058	matrix metallopeptidase-1, 2, 3, 9, 11, 13, meprin alpha subunit, cathepsin B, D, E, L, S
Ile Met	0,0058	elastase-2, matrix metallopeptidase-2

Terminal C Residue	Clivages Ratio	Peptidases
Ile Ala	0,0191	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Gln Gly	0,0133	Cathepsin L,S,K
Lys Ala	0,0133	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Tyr Gly	0,0106	Pepsina A; Cathepsin D,L,S,B
Leu Ala	0,0090	Pepsin A, Cathepsin D and E
Ile Thr	0,0074	pepsin A, Cathepsin E, L and G, rhizopuspepsin, elastase-2, ERAP2 aminopeptidase
Val Gly	0,0048	matrix metallopeptidase-9 and 12, elastase-2, HtrA2 peptidase, rhizopuspepsin
Arg Gly	0,0048	rhizopuspepsin, memapsin-1, yapsin A, cathepsin B, L and K

Uvula and Soft Palate SCC

Terminal N Residue	Clivages Ratio	Peptidases
Ile Met	0,0164	elastase-2, matrix metallopeptidase-2
Ile Ala	0,0111	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Tyr Val	0,0095	pepsin A, rhizopuspepsin, cathepsin D and E, HIV-1 retropepsin
Arg Ser	0,0069	yapsin-1 and 3, cathepsin B, L, S, K, P, rhodesain, gingipain R
Asn Ser	0,0058	aspergillopepsin, cathepsin, L, S and B
Val His	0,0058	gastricsin, pepsin A, falcipain-2, elastase-2, insulysin, cathepsin G
Met Gly	0,0037	aspergillopepsin I, cathepsin B, L, S, K,
Ala Met	0,0037	matrix metallopeptidase-1, 2, 3, 9, 11, 13, meprin alpha subunit, cathepsin B, D, E, L, S
Tyr Ser	0,0037	gastricsin, aspergilloglutamic peptidase, calpain-2, chymotrypsin A, cathepsin E, L and S
Phe His	0,0032	phytepsin, cathepsin B, L and S, chymotrypsin A, carboxypeptidase A6, matrix metallopeptidase-2, 8 and 13

Terminal C Residue	Clivages Ratio	Peptidases
Ile Ala	0,0191	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Lys Ala	0,0133	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Ile Thr	0,0133	pepsin A, Cathepsin E, L and G, rhizopuspepsin, elastase-2, ERAP2 aminopeptidase
Val His	0,0106	gastricsin, pepsin A, falcipain-2, elastase-2, insulysin, cathepsin G
Val Tyr	0,0090	falcipain2 and 3, matrix metallopeptidase 2 and 13, HtrA2 peptidase, elastase-2
Phe Tyr	0,0074	pepsin A and B, chymosin, cathepsin D and E
Ala His	0,0058	necepsin-2, Cathepsin L and S, falcipain 2 and 3
Val Leu	0,0048	Pepsin A, Cathepsin D and E, rhizopuspepsin, necepsin 1 and 2
Tyr Gly	0,0048	Pepsina A; Cathepsin D,L,S,B
Lys Ser	0,0042	pepsin A, cathepsin D, E and L, necepsin-1, calpain 1 and 2

Tonsils SCC

Terminal N Residue	Clivages Ratio	Peptidases
Gln Gly	0,0329	Cathepsin L,S,K
Gln Ser	0,0101	pepsin A, gastricsin, Cathepsin B, L and S
Tyr Val	0,0074	pepsin A, rhizopuspepsin, cathepsin D and E, HIV-1 retropepsin
Val Leu	0,0042	Pepsin A, Cathepsin D and E, rhizopuspepsin, necepsin 1 and 2
Arg Lys	0,0037	pepsin A, Cathepsin D, E and L, omptin, peptidyl-Lys metallopeptidase
Arg Pro	0,0037	bacilliform virus peptidase, aminopeptidase Ey, P1 and P2, trypsin 1
Asp Leu	0,0037	pepsin A, cathepsin D and E, caspase 1, 3, 6, 7 and 8
Leu Asp	0,0037	pepsin A, cathepsin D, E and S, necepsin-1, meprin beta subunit
Pro Arg	0,0032	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Lys Gly	0,0032	pepsin A, rhizopuspepsin, cathepsin L and S, falcipain 2 and 3, gingipain K, matrix metallopeptidase 2, 9 and 13
Ile Arg	0,0032	pepsin A, rhizopuspepsin, cathepsin L, calpain-1 and 2, elastase-2

Terminal C Residue	Clivages Ratio	Peptidases
Gln Gly	0,0393	Cathepsin L,S,K
Leu Thr	0,0111	pepsin A and F, cathepsin D, E L and S, phytepsin, Rous sarcoma virus retropepsin, chymotrypsin A, granzyme A and M
Lys Ala	0,0074	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Val Tyr	0,0064	falcipain2 and 3, matrix metallopeptidase 2 and 13, HtrA2 peptidase, elastase-2
Val Leu	0,0058	Pepsin A, Cathepsin D and E, rhizopuspepsin, necepsin 1 and 2
Pro Arg	0,0058	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Tyr Gly	0,0048	Pepsina A; Cathepsin D,L,S,B
Gln Ser	0,0042	pepsin A, gastricsin, Cathepsin B, L and S
Ser Arg	0,0042	myeloblastin, pepsin A, gastricsin, memapsin-1,

Oropharyngeal SCC

Terminal N Residue	Clivages Ratio	Peptidases
Val Arg	0,0398	DegP peptidase; Cathepsin L; chymosin;rhizopuspepsin
Met Asn	0,0350	granzyme M, meprin beta subunit, Cathepsin B, E, L, K, S
Lys Ala	0,0308	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Ile Met	0,0239	elastase-2, matrix metallopeptidase-2
Ile Ala	0,0143	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Tyr Val	0,0143	pepsin A, rhizopuspepsin, cathepsin D and E, HIV-1 retropepsin
Leu Ala	0,0138	Pepsin A, Cathepsin D and E
Tyr Gly	0,0133	Pepsina A; Cathepsin D,L,S,B
Phe Gly	0,0106	pepsin A and B, cathepsin D, E and L, necepsin 1, HIV-1 retropepsin, chymotrypsin A
Val Leu	0,0085	Pepsin A, Cathepsin D and E, rhizopuspepsin, necepsin 1 and 2

Terminal C Residue	Clivages Ratio	Peptidases
Ile Ala	0,0679	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Lys Ala	0,0377	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Lys Ser	0,0175	pepsin A, cathepsin D, E and L, necepsin-1, calpain 1 and 2
Tyr Gly	0,0164	Pepsina A; Cathepsin D,L,S,B
Ala His	0,0164	necepsin-2, Cathepsin L and S, falcipain 2 and 3
Ala Gly	0,0143	cathepsin L, S, K and B
Val Leu	0,0133	Pepsin A, Cathepsin D and E, rhizopuspepsin, necepsin 1 and 2
Tyr Thr	0,0133	SplA peptidase ({Staphylococcus aureus}), chymotrypsin A, meprin alpha subunit, rhizopuspepsin,
Arg Ile	0,0111	matrix metallopeptidase 2, 12 and 13, kallikrein-related peptidase 5, 12 and 14, cathepsin E, P and H

Nasopharynx

Terminal N Residue	Clivages Ratio	Peptidases
Ile Ala	0,0111	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Gln Gly	0,0095	Cathepsin L,S,K
Tyr Gly	0,0085	Pepsina A; Cathepsin D,L,S,B
Tyr Val	0,0074	pepsin A, rhizopuspepsin, cathepsin D and E, HIV-1 retropepsin
Tyr Ser	0,0074	gastricsin, aspergilloglutamic peptidase, calpain-2, chymotrypsin A, cathepsin E, L and S
Leu Ala	0,0064	Pepsin A, Cathepsin D and E
Cys Val	0,0048	cathepsin B, E and L, necepsin 1 and 2, thermolysin, bacillolysin, caspase-1
Ala Ala	0,0048	pepsin A, cathepsin B, E, L, S and K, matrix metallopeptidase 2, 3, 7, 9, 11 and 12
Lys Ala	0,0048	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Lys Leu	0,0048	falcipain 2 and 3, calpain 1 and 2, bacillolysin, trypsin 1, membrane-type matrix metallopeptidase 1, 3 ,4 and 6, matrix metallopeptidase2, 3, 7, 9 and 12

Terminal C Residue	Clivages Ratio	Peptidases
Lys Ala	0,0164	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Lys Ser	0,0122	pepsin A, cathepsin D, E and L, necepsin-1, calpain 1 and 2
Ile Ala	0,0111	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Gln Gly	0,0064	Cathepsin L,S,K
Tyr Gly	0,0053	Pepsina A; Cathepsin D,L,S,B
Pro Arg	0,0053	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Ile Thr	0,0048	pepsin A, Cathepsin E, L and G, rhizopuspepsin, elastase-2, ERAP2 aminopeptidase
Ala Tyr	0,0042	cathepsin B, D and E, plasmepsin-1, pepsin A, matrix metallopeptidase 2, 3, 7, 9 and 12

Hypopharyngeal SCC

Terminal N Residue	Clivages Ratio	Peptidases
Gln Gly	0,0202	Cathepsin L,S,K
Ser Gln	0,0048	calpain 1 and 2, Cathepsin L and , matrix metallopeptidase-2
Pro Pro	0,0048	Cathepsin L1 and L3, lactocepin I, calpain 1 and 2
Gln Pro	0,0032	lactocepin I and 3, elastase-2, thermolysin
Arg Gly	0,0027	rhizopuspepsin, memapsin-1, yapsin A, cathepsin B, L and K
Lys His	0,0021	rhizopuspepsin, syncephapepsin, aspergillopepsin I, trypsin 1, lysyl endopeptidase (bacteria)
Phe His	0,0021	phytepsin, cathepsin B, L and S, chymotrypsin A, carboxypeptidase A6, matrix metallopeptidase-2, 8 and 13
Arg Lys	0,0016	pepsin A, Cathepsin D, E and L, omptin, peptidyl-Lys metallopeptidase
Arg Ser	0,0016	yapsin-1 and 3, cathepsin B, L, S, K, P, rhodesain, gingipain R
Pro Arg	0,0016	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Gly Arg	0,0016	necepsin-1, cathepsin B, S, L and X, carboxypeptidase B, E, M, N, and U
Tyr Gly	0,0016	Pepsina A; Cathepsin D,L,S,B
Phe Gly	0,0016	pepsin A and B, cathepsin D, E and L, necepsin 1, HIV-1 retropepsin, chymotrypsin A

Terminal C Residue	Clivages Ratio	Peptidases
Gln Gly	0,0186	Cathepsin L,S,K
Tyr Gly	0,0069	Pepsina A; Cathepsin D,L,S,B
Arg Gly	0,0032	rhizopuspepsin, memapsin-1, yapsin A, cathepsin B, L and K
Pro Arg	0,0027	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Ser Arg	0,0027	myeloblastin, pepsin A, gastricsin, memapsin-1,
Pro Gln	0,0027	ADAM8 peptidase, ADAM10 peptidase, ADAM19 peptidase, prolyl oligopeptidase, membrane-type matrix metallopeptidase 1, 2, 3, 4 and 6, matrix metallopeptidase 2, 8, 9 and 12
Gln Pro	0,0027	lactocepin I and 3, elastase-2, thermolysin
Gln Tyr	0,0027	Cathepsin E, L, S and B, matrix metallopeptidase-2

Nostrils

Terminal N Residue	Clivages Ratio	Peptidases
Tyr Val	0,0117	pepsin A, rhizopuspepsin, cathepsin D and E, HIV-1 retropepsin
Ile Met	0,0117	elastase-2, matrix metallopeptidase-2
Gln Gly	0,0090	Cathepsin L,S,K
Tyr Gly	0,0069	Pepsina A; Cathepsin D,L,S,B
Asp Gly	0,0058	pepsin A; Cathepsin B, Caspase 1, HIV-1 retropepsin
Met Asn	0,0058	granzyme M, meprin beta subunit, Cathepsin B, E, L, K, S
Met Gly	0,0058	aspergillopepsin I, cathepsin B, L, S, K,
Ile Ala	0,0058	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Lys Ala	0,0058	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Val His	0,0048	gastricsin, pepsin A, falcipain-2, elastase-2, insulysin, cathepsin G

Terminal C Residue	Clivages Ratio	Peptidases
Ile Ala	0,0233	Pepsin A, Cathepsin E, L, S; rhizopuspepsin,aspergillopepsin I
Pro Arg	0,0122	memapsin-1, angiotensin-converting enzyme-2, matrix metallopeptidase 2 and 9
Lys Ala	0,0122	Pepsin A, penicillopepsin, rhizopuspepsin, aspergillopepsin I, cathepsin L and X
Leu Ala	0,0080	Pepsin A, Cathepsin D and E
Gln Gly	0,0074	Cathepsin L,S,K
Tyr Gly	0,0064	Pepsina A; Cathepsin D,L,S,B
Ile Thr	0,0058	pepsin A, Cathepsin E, L and G, rhizopuspepsin, elastase-2, ERAP2 aminopeptidase
Val Tyr	0,0048	falcipain2 and 3, matrix metallopeptidase 2 and 13, HtrA2 peptidase, elastase-2
Arg Asp	0,0042	granzyme A, trypsin 1, meprin alpha subunit, meprin beta subunit

Number of peptide sequences for each individue

Control	nº of repeated sequences	nº of unique sequences	nº of total sequences
C1	52	79	131
C2	42	99	141
C3	27	60	87
C4	23	55	78
C5	29	106	135
C6	5	34	39
C7	2	28	30
C8	42	70	112
C9	8	68	76
C10	18	39	57
Average	**24,8**	**63,8**	**88,6**
STDev	16,9	26,2	40,0

Vocal Chords (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
1	11	43	54
2	100	77	177
3	35	51	86
4	11	70	81
Average	**39,2**	**60,2**	**99,5**
STDev	42,0	15,9	53,5

Larynx (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
5	18	108	126
6	23	71	94
7	76	83	159
8	46	39	85
9	4	74	78
10	43	76	119
11	10	69	79
12	40	36	76

13	73	71	144
14	36	84	120
15	21	15	36
16	49	156	205
17	49	25	74
18	6	28	34
19	35	107	142
20	10	61	71
21	31	58	89
22	11	54	65
23	10	30	40
24	11	50	61
Average	**30,1**	**64,75**	**94,85**
STDev	21,4	33,6	44,3

Tongue (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
25	49	110	149

Oropharynx (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
26	89	108	197
27	3	49	52
28	60	46	106
29	144	140	284
Average	**74**	**85,75**	**159,75**
STDev	58,7	46,0	102,1

Mouth floor (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
30	78	100	178
31	17	114	131
Average	**47,5**	**107**	**154,5**
STDev	43,1	9,8	33,2

Uvula and Soft Palate (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
32	**66**	**44**	**110**

Tonsil (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
33	51	123	174
34	5	40	45
35	40	34	74
Average	**32**	**65,6**	**97,6**
STDev	24,03	49,7	67,64

Pharynx (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
36	13	97	110

Pharyngolarynx (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
37	95	160	255
38	81	167	248
Average	88	163,5	251,5
STDev	9,8	4,9	4,9

Hypopharynx (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
39	2	6	8
40	29	99	128
Average	15,5	52,5	68
STDev	19,0	65,7	84,8

Nostrils (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
41	69	313	382
42	8	44	52
43	51	18	69
Average	42,6	125	167,6
STDev	31,3	163,3	185,8

Nasopharynx (patient)	nº of repeated sequences	nº of unique sequences	nº of total sequences
44	66	95	161

Control n=10	repeated sequences	unique sequences	total sequences
Average	24,8	63,8	88,6
STDev	16,9	26,2	40,01

Larynx group n=24	repeated sequences	unique sequences	total sequences
Average	31,6	64	95,6
STDev	24,9	31,1	44,7

Pharynx and Oral Cavity Group n=16	repeated sequences	unique sequences	total sequences
Average	51,375	89,8125	140,5625
STDev	39,9	47,8	78,8

Nasopharynx and Nostrils n=4	nº of repeated sequences	nº of unique sequences	nº of total sequences
Average	48,5	117,5	166
STDev	28,1	134,2	151,7