Mémoire de Maîtrise en médecine No 1951

Molecular determinants of HPV-induced oropharyngeal squamous cell carcinoma

Etudiante
Gwaenaelle Aellen

Tuteur
Prof. Christian Simon
Dpt de chirurgie ORL, CHUV

Co-tuteur
Dr Genrich Tolstonog
Dpt de recherche chirurgie ORL, CHUV

Expert
Dr Marie-Catherine Vozenin
Dpt de radio-oncologie, CHUV

Lausanne, 15.01.2015

UNIL | Université de Lausanne
Faculté de biologie et de médecine

CHUV Centre hospitalier universitaire vaudois

Contents

1. Introduction

1.1. Head and Neck Squamous Cell Carcinoma (HNSCC)

Head and neck cancers are the six most malignant tumors in the world. There are 600,000 new cases diagnosed worldwide each year(1). It may arise anywhere in the oral cavity, nasopharynx, oropharynx, hypopharynx and larynx. Tumors in nasal cavities, paranasal and sinuses are also included (Figure 1). Despite improvements in treatment strategies survival rate is still low for the advanced cases; the 5-years survival at early stage is about 80% and drop to 19% for late stage. Around 95% of head and neck cancers are squamous cell carcinoma and frequently have aggressive biological behavior.

Etiology of head and neck cancers is multifactorial, but it is closely linked with the environmental and lifestyle risk factors including chronic tobacco smoking (strongly associated), alcohol consumption, actinic radiation and inherited genomic instability. Typical HNSCC patients are middle-aged with a history of chronicle use of alcohol and tobacco(1).

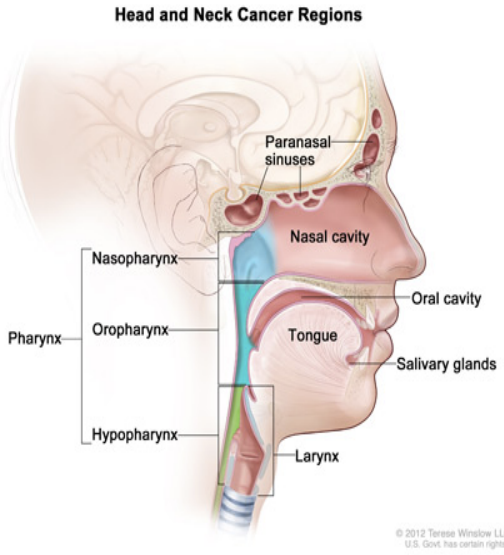

Figure 1: Head and neck cancer localizations.

Regions of some anatomic structures that are targeted in HNSCC cancer.

Paranasal sinuses, nasal cavity, oral cavity, tongue, salivary glands, larynx and pharynx (including nasopharynx, oropharynx and hypopharynx).

Picture taken from web page: http://www.cancer.gov/PublishedContent/Images/cancertopics/factsheet/Sites-Types/headandneck-diagram.jpg

Tobacco smoking and alcohol consumption–associated head and neck cancer is developing by a multistep progression and accumulation of genetic and epigenetic alterations, including the activation of oncogenes and inactivation of tumor-suppressor genes. Typical alterations in genes implicated in HNSCC carcinogenesis are loss of heterozygosity and hypermethylation particularly at chromosomes 3p, 9p and 17p, resulting in the inactivation of p16 (inhibitor of cyclin-dependent kinase). These alterations are typical for early stage of carcinogenesis found in dysplasia. In addition, alterations at the chromosomes 11q, 4q and

chromosome 8 are typical for late stage carcinogenesis(1,2). TP53 mutations are present in 60%-80% of all head and neck cancers. Those mutations are hallmarks of HNSCC carcinogenesis(1,2). Others mutations like EGFR mutations may also be present but they are not frequent.

The prognosis for patients with HNSCC is largely determined by the stage of the cancer defined as “staging”. Treatments mainly depend on the tumor localization. Tumor size (T), lymph nodes extension (N), metastasis (M) stages are additionally used to determine the treatment. Indeed, TNM stages are used to classify different groups in early, loco-regionally or metastatic stages to determine the treatment. Treatments traditionally involved are chemotherapy, radiotherapy and surgery according to the staging(1). Other additional treatments like immunotherapy, which targets specific pathways (i.e. cetuximab for EGFR pathway), can also be employed(2).

As mentioned above, HNSCC is strongly associated with chronic tobacco smoking. Between 1950 and 1960, studies demonstrated the evidence of the association between cigarettes - lung cancer and between cigarettes - cardiovascular disease. Those evidences were proved to diminish tobacco used in population. According to the robust correlation between HNSCC and tobacco, it was expected the incidence of HNSCC to decrease a few decades after. Indeed, statistics show a decrement of cancers localized in oral cavity, pharynx and floor of the mouth. On the other hand, incidence of cancers localized in tongue, tonsils and oropharynx increased dramatically(2), suggesting another carcinogens exposure emerging in oncogenesis process of head and neck cancer.

For a long time HNSCC has been considered like one monotonous disease entity, but recent studies highlighted a subset of HNSCC related to Human Papillomavirus (HPV). Indeed, some studies show that 20% of HNSCC are HPV positive. Furthermore, HPV are 70% causative of oropharyngeal cancer including tonsils, base of tongue, soft palate and pharyngeal wall.
It is clear now that HPV affects the demographical, clinical and histopathological presentation of the HNSCC patient and should be considered as a unique clinical entity.

1.2. Head and neck cancer associated with Human Papilloma Virus (HPV)

HPV is a sexually transmitted virus, especially known for its great implication in cervix carcinoma amongst female population. It is also responsible for penile cancer and rectal cancer more prevalent amongst male population. HPV is recently known to have also an implication in HNSCC. Indeed, HPV plays a role in the pathogenesis of a subset from these cancers, particularly in those that arise in the area of oropharynx (OPC) and in the lingual and palatine tonsils(3)**.**

The etiology of HNSCC linked to HPV infection is predominantly sexually transmitted. In fact, the major risk factors for HPV-positive HNSCC are certain sexual behaviors, such as high number of sexual partners, history of performing oral sex and oral-anal sex, history of genital warts, and infrequent use of condoms. Other factors are more linked with the immune system status, including aging, history of sexual transmitted disease, HIV infection, severity in immunosuppression (transplant patients), and Fanconia anemia (genetic susceptibility)(3–5).

In most cases, HPV-positive oropharyngeal cancer (OPC) is associated with high-risk HPV 16 subtype. Indeed, high-risk HPV 16 according to some studies seems to be causally implicated in about 90-95% of HPV-positive HNSCC tumors(3,4). Other subtypes including HPV 18, -31, -33, -45 and -53 are also important etiological agents.

Prevalence of HPV varies widely depending on geographical location and on sexual behaviors. In 2009, prevalence of HPV type 16 genital infections in the United States was about 27% amongst women and about 60% amongst men (both gender aged 14 to 59 years). As mentioned before, the prevalence of classical HNSCC (associated with risk factors including tobacco and alcohol) is decreasing whereas the prevalence of HNSCC linked to HPV was constantly increasing last decade and continues to increase nowadays (5).

Another study in 2002 estimated that about 5% of worldwide new cancer cases were attribute to HPV, making HPV one of the most important infectious causes of cancers (6).

HPV totally changed the face of HNSCC traditionally associated with tobacco use. Indeed, patients HPV-positive HNSCC tend to be younger (4). Moreover, HPV status seems to be an important prognosis factor. Previous studies show that HPV status is associated with favorable outcomes, better overall survival, disease-free survival and lower recurrence rates(5). It has also been demonstrated that this status has better response to treatments, in particularly with induction chemotherapy, radiotherapy and chemoradiotherapy (7). Patients

with HPV-positive status in HNSCC present less exposure to tobacco and chronicle use of alcohol. This lack of exposure could be one explanation among others of the better prognosis.

1.3. Human Papilloma Virus 16 (HPV 16)

HPV is a small circular double-stranded DNA virus of approximately 8000 base pairs, and has a specific tropism for squamous cell epithelia (basal keratinocytes)(8). More than 120 different types of HPV have been identified and separated in two groups according to their tissue tropism and their carcinogenesis ability: low-risk and high-risk. Low-risk HPV (e.g. HPV-6, -11) are cutaneous types found in benign genital warts and in other non-malignant lesions. In contrary, high-risk HPV (e.g. HPV-16, -18) are mucosal types found in precancerous and cancerous lesions.

High risk HPV16 is the most virulent and aggressive HPVs and is the main cause of cervix cancer and of a subset of HNSCC. Its genome consists of: non-coding region, 6 early genes (E1-7) and 2 late genes (L1-2). Other regions such as long control region (LCR) are also present in the genome's virus, which are important for the production of the virus (8) (Figure 2).

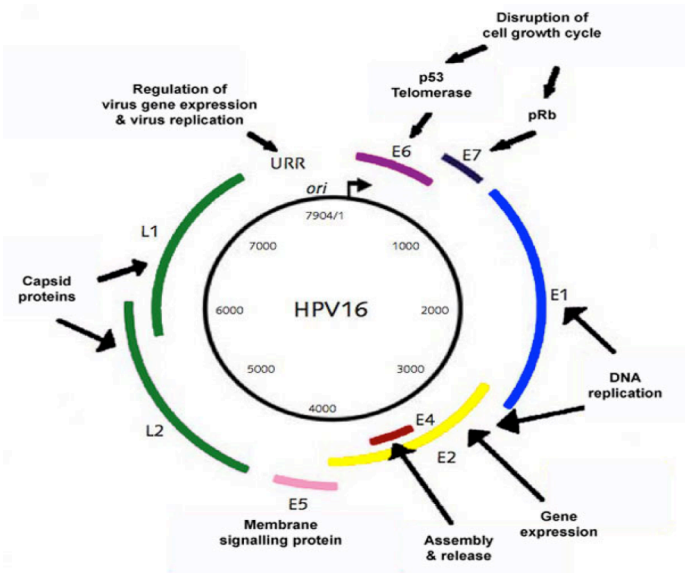

Figure 2: HPV 16 model.

1) Early region (E1-7) codes for non-structural genes (cellular transformation).

2) Late region (L) codes for two structural proteins L1 and L2 that form the capsid.

3) Long Control Region (LCR), a non-coding region that regulates replication and gene functions.

Oncoproteins E6 and E7 are required for tumor maintenance.
E6 inhibits p53 (stop apoptosis), E7 inhibits Rb (cell growth arrest).

Picture from the net: http://www.microbiologybytes.com/virology/Papillomaviruses.html

E6 and E7 are two oncogenes that are the onset of HPV mediated carcinogenesis(9). They block some important tumor suppressors and activate oncogenes that lead to cell cycle

deregulation and cell immortalization. The loss of this regulation leads to neoplasia. Accumulation of others genetic damages will ultimately lead to malignant transformation. All these steps will be described in the next paragraph.

1.4. Invasion and carcinogenesis of high-risk HPV

HPV associated cancer in oropharynx might share many features with carcinogenesis in cervical cancer amongst women.

Carcinogenesis of HPV 16 begins with basal cell infection of the epithelium. This infection is happening only when there is abrasion (micro wounds) in the epithelia. The virus reaches the basal layer of the epithelia and infects some basal cells (Figure 3)(10). HPV16 depends on the host cell replication machinery to copy its own DNA. To integrate DNA host cell the necessary proteins are only produced during the mitosis (actively dividing cell)(11).

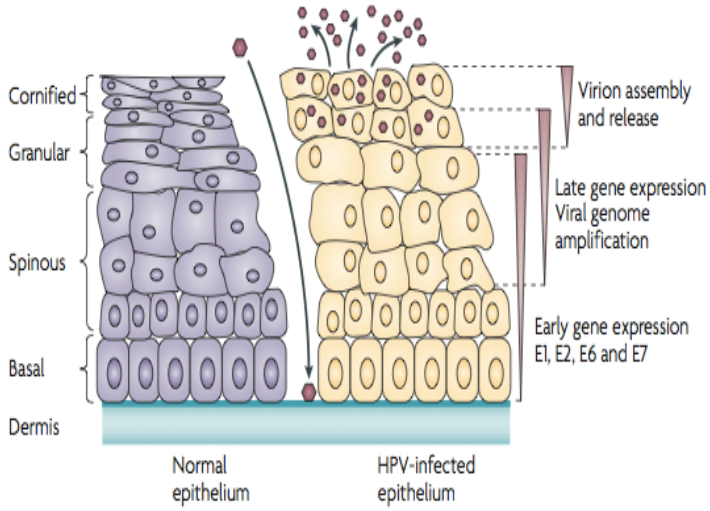

Figure 3: Natural way of HPV infection

HPV infects keratinocytes in the basal layer that are exposed by microwounds. Differentiation of HPV cells induces the productive phase of the viral life cycle, which requires cellular DNA synthesis machinery of the host cell. The expression of E6 and E7 deregulates cell cycle control that pushes the cell in the S phase that allows viral replication and amplification.

Picture from NATURE REVIEW CANCER, Human papilloma oncoproteins: pathway to transformation, C Moody et All.

Previously, studies have shown that mitosis phase progression is required for early gene expression of HPV. Resulting population of host cells with the integrated HPV DNA into their genomes will accumulate genetic damages and alterations. These transformations in the important cell cycle pathways can ultimately lead to malignant transformation and to carcinoma(11).

Cellular transformation and maintenance of malignant phenotype is accomplished by high-risk HPV 16 oncoproteins (E6 and E7). There are other oncoproteins, which can play a role in the carcinogenesis of HPV, but they will not be discussed in this study(12).

The carcinogenesis and the maintenance of malignancy are driven by the E6 and E7 proteins and it is shown in the Figure 4.

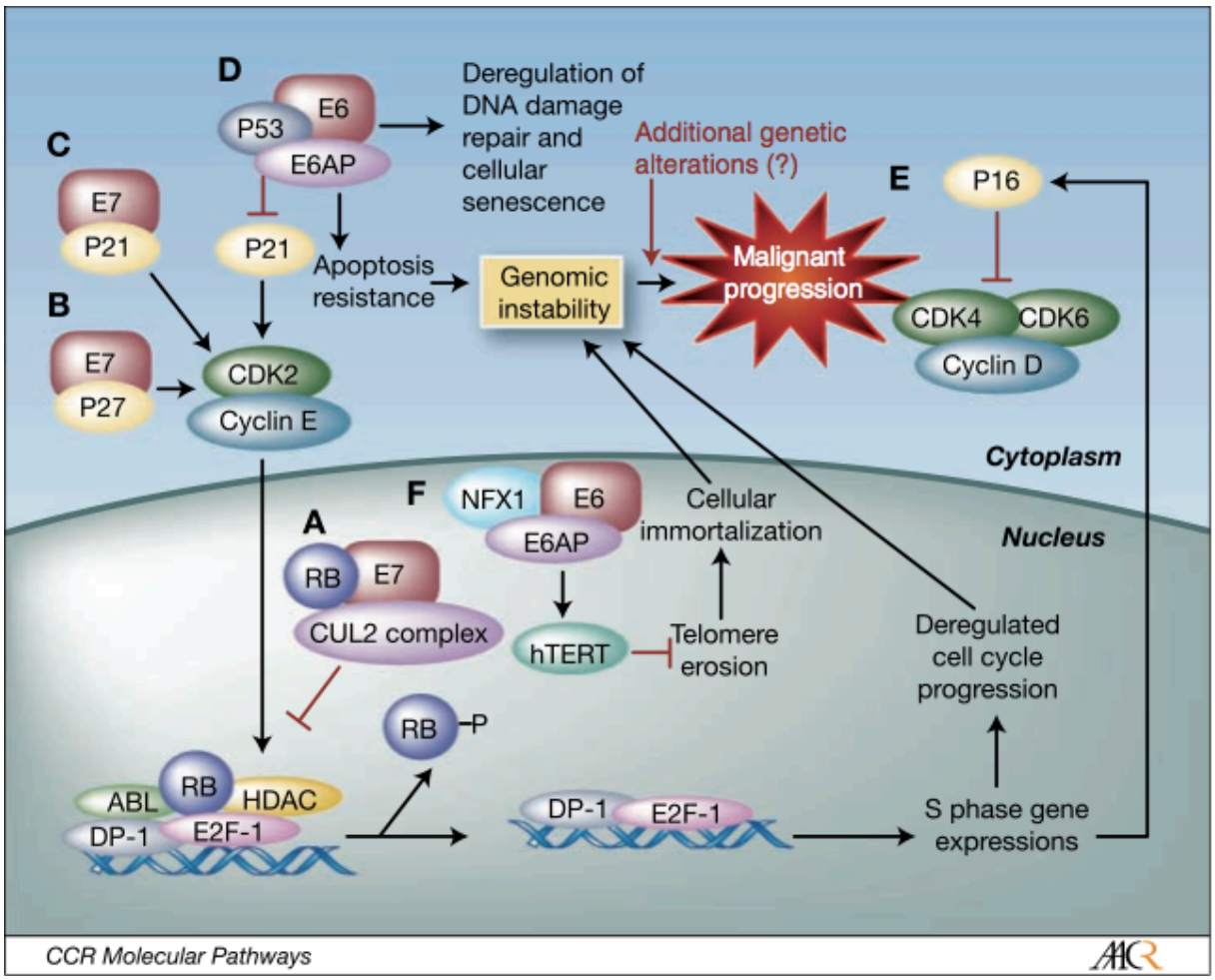

Figure 4: carcinogenesis of E6 and E7, *picture taken from Clinical Cancer Research, Human Papilloma Virus in Head and Neck Cancer: its Role in Pathogenesis and Clinical Implications, H. Chung et all, 2009.*

E6 is required to deregulate both G1/S and G2/M cell cycle checkpoint by interaction with E6-associated protein (E6-AP) and targeting of p53 tumor suppressor protein for ubiquitination, which results in rapid degradation of p53. Inactivation of p53 results in apoptosis resistance and accumulation of DNA damages which lead to genomic instability. The protein complex E6/E6-AP plays also a role in cellular immortalization by degrading NFX1 that is a transcriptional repressor of hTERT that results in its activation and in cellular immortalization by inhibiting telomerase erosion(12,13).

E7 is required to deregulate G1/S phase of the cell cycle by binding cullin 2 ubiquitin ligase which ubiquitinates the retinoblastoma (pRb) tumor suppressor protein. The absence of pRb function leads to cell proliferation because E2F family of transcription factors can induce transcription of the S phase genes when pRb is not functional. In addition, p16 is a cyclin-

dependent kinase inhibitor whose expression increases in response to Rb inactivation(12,14). Therefore the expression status of p16 is strongly correlated with tumor HPV status(5,15). So it is commonly used as a surrogate marker in HPV-positive tumors detection. Indeed, most of tumors (96%) that are positive for HPV by In Situ Hybridization are also positive for p16 Immunohistochemistry(4,15). It is now well known that E6 and E7 are required for tumor maintenance in oropharyngeal cancer as in the cervix cancer, because studies have demonstrated when E6 and E7 are inhibited in a tumor cell it induces apoptosis and decrease cell viability(16).

The process of transforming normal cell into cancerous ones is very slow; therefore HPV-induced tumors may occur in people that have been infected with HPV for a long time ago (10-15 years). It is important to notice that majority of individuals infected will not develop cancer, because most of the infections are rapidly cleared by the immune system and do not progress to carcinoma. In all steps of the carcinogenic process there is a possibility for the immune system to clear HPV and stops the progression to carcinoma(11,16).

1.5. Prognosis and response to treatment of HPV-positive oropharyngeal cancer (OPC)

Patients with HPV-positive OPC tend to be younger by 5-10 years and are more likely to be non-smokers and non-drinkers compared to HPV-negative OPC patients.

There is evidence that HPV is an important prognosis factor associated with better outcomes and survival compared to patients with HNSCC HPV-negative tumors. In fact retrospective studies demonstrated a 50-80% reduction risk of cancer related death for HPV-positive than HPV-negative tumors. Many others studies have shown that after adjustment of age and tumor stages, HPV-positive tumors are associated with better overall survival, local recurrences free survival, disease-free survival and better response to treatment (7).

Improved prognosis and response to therapy in HPV-positive tumors may be associated with the absence of TP53 mutations. In HPV transformed cells p53 and Rb tumor suppressor genes are wild type, meaning that the pathways are present and intact but dormant due to the expression of E6 and E7 proteins. In the absence of mutations, the apoptotic pathways still functional and can be partially reactivated in response to radiation treatment compared

with the HNSCC HPV-negative tumors in which mutations are limiting the response to therapy.

Smocking seems to affect the survival of patients with HPV-positive OPC in a negative way. Studies demonstrated that overall survival and recurrences were better in HPV-positive OPC non-smokers that in HPV-positive OPC smokers. As consequence, tobacco has negative influences even in patients with HPV-positive tumor. We could explain these negative influences by the fact that smoking may contribute to increase in EGFR expression (17).

1.6. HPV Vaccine

Some vaccines are available on the market and are a possible intervention to fight against HPV, in particularly Gardasil® that targets most prevalent HPV-6, -11, -16, and -18 subtypes. This vaccine has been demonstrated to be highly effective in preventing cervix cancer amongst female. If it is successful by reducing the prevalence of HPV16, we would expect that it will also reduce the prevalence of OPC HPV-positive tumors, but it remains to be determined. This expectation can be used as an argument to extend vaccination program and include boys for the future (18). However, prophylactic HPV vaccines will eventually reduce infections by the most prevalent high-risk HPV subtypes, but do not cover all high risks HPVs. Furthermore, this vaccine has demonstrated to have a lack of therapeutic action, meaning that it will not affect the already existing HPV infections. Those infections could lead to a probable cancer in a decade or more. Taking that into consideration, it encourages to broaden an interest of research for HPV tumors.

1.7. Aim of the study

HPV-positive HNSCC demonstrated to have completely different clinical, histological and molecular features and also showed to have a better prognostic and response to treatment than the HPV-negative HNSCC.

However, patients with HPV-positive tumors continue to be treated exactly the same way as the HPV-negative patient (19). A standardization of assay methodology is required to reliably detect HPV-positive tumor either by a direct detection of virus related molecules (DNA, RNA

and proteins) or detection of the specific molecular markers related to the HPV-induced tumorigenesis. It has to be sensitive, cost effective and easy to use in laboratories (20).

The aim of this study is first to test some biopsy samples to detect HPV in tumor, and secondly to measure expression of genes related to the HPV-induced tumorigenesis in order to find a possible marker (21,22).

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Reagents

PCR reaction mixture:

12.5 µl of Go Taq DNA polymerase

0.6 µl of Forward and Reverse primers

2 µl of template (cDNA)

Adjusted to 25 µl with ddH_20

qRT-PCR reaction mixture:

5 µl of Fast SYBR Green Master mix 2x

0.2 µl of Forward and Reverse primers 10µM stock

2 µl of template (cDNA)

2.1.2. Equipment

-Plastic vessels (VWR International GMBH)

-Biometra TProfessional Basic Thermal Cycler

- Thin-walled and flat-capped tube (Chemie Brunschwig AG)

-Go Taq DNA polymerase master mix (Promega; contains GoTaq® DNA Polymerase, 5X Green GoTaq® Reaction Buffer and 5X Colorless GoTaq® Reaction Buffer)

-Agarose gels (Chemie Brunschwig AG)

-Material for PCR: combs (BioRad)

- Inverted Zeiss microscope

-Zymo kit and Direct-ZOL RNA Mini-Prep (Zymoresearch)

-BCA protein assay kit (Pierce # 23227)

-Blotek Epoch system

-Gens 1.10 Software

- ClustalW software

- Clone Manager Professional software (Version 9 for Windows, http://www.scied.com)

-Mini-Protean II dual slab cells (Bio-Rad)

-Mini Protean II transfer system (Biorad)

-Nitrocellulose sheet (Protran BA-85, 0.45 μ m)

- Luminol-based ECL Western blotting detection reagents from Pierce (SuperSignal West Pico)

-Falcon tube

-Mixer plate

-Vortex

-Centrifuge

- Nanodrop spechtrophotometry (test type: nucleic acid)

-RX detection box

- X-ray film (Fuji)

2.1.3. Buffers and solutions

10x running buffer (SDS-PAGE)

Trizma base: 30.3 g

Glycine: 144 g

20% SDS: 50 ML

Add H2O to reach 1 litter

10x transfer buffer (western blotting)

Trizma base: 30.3 g

Glycine: 144 g

10% Ethanol or 20% methanol for small proteins

Add H2O to reach 1 litter

KLB-Buffer

25 mM Tris-HCl (pH 7.4)

150 mM NaCl,

5 mM EDTA

10% glycerol,

1% Triton X-100,

10 mM sodium pyrophosphate,

10 mM beta-glycerol phosphate,

1 mM sodium orthovanadate,

1 mM DTT,

1 mM PMSF,

0.02 mg/ml aprotinin,

10 mM sodium fluoride

0.1 mM sodium pervanadate

2.1.4. Cell lines and antibodies

Cell lines	Localization	HPV Status	P53 Status
SCC-090	Oral cavity	HPV 16 positive	Wild-type p53
LAU 2068	Oral cavity	-	Mutant p53
LAU 2106	Oral cavity	-	Mutant p53
LAU 2092	Oropharynx	-	Wild-type p53
LAU 2089	Oral cavity	-	Wild-type p53

Tumour samples	Localization	HPV Status
25	Oral cavity	HPV 33 positive
27	Oral cavity	HPV negative

Antibodies	Species	Dilution
SHISA2	Rabbit polyclonal	1:2500

2.2. Methods

2.2.1. Tissue samples

RNA from 40 biopsies was from the University Hospital of the canton de Vaud (CHUV). All tissue samples were collected during the surgery and before the initiation of any previous treatment (chemotherapy or radiotherapy). The biopsies were coming from oral cavity and oropharynx. All samples were collected with patient consent under approval of the ethics commission of the institution (protocol 89/08).

Cell lines

Human SCC-090 cell line (HPV16-positive) was purchased from DSMZ (Germany); Hela (ATCC CCL-2™) cell line was purchased from ATCC, TC-1 cell line was kindly provided by Dr. Denise Nardelli Haefliger (CHUV). LAU-2068, LAU-2081, LAU-2089, LAU-2092, and LAU-2106 cell lines were established by Dr. Jean-Paul Rivals (CHUV) from head and neck cancer samples.

Cell culture methods

The preparation of media and all cell culture work were performed in a laminar flow hood, using standard techniques (Freshney, 1987). Cells were grown in a CO_2 incubator in an atmosphere containing 5% CO_2 either attached to plastic (anchorage-dependent) or in suspension in sterile, disposable plastic vessels (VWR International GMBH). Diluting the cultures with fresh medium every 2-3 days carried out feeding and subculturing of suspension cells. Passage of adherent cells was performed every 3-10 days, when they had reached confluency. Most cell lines reached confluence after 3-4 days, however, SCC-090 cell line grow slow reaching confluence after around 10 days. The cells were then detached from the dish with trypsin, terminating the proteolysis reaction with FCS. We checked the detachment of the cells with an inverted Zeiss microscope. The cells were collected by centrifugation (400xg, 5 min, at room temperature) and plated in fresh medium. When extracts had to be prepared, the cells were mechanically scraped off from the substratum with a rubber policeman in a lysis buffer.

Viable cell count

A small sample (0.1 ml) of suspension culture was mixed with the same volume of a solution prepared from 4 parts of a 0.2% (w/v) Trypan blue stock solution and 1 part of 5X saline (4.25% NaCl in dist. H_2O). The total number of cells and the number of stained cells were counted in a hemocytometer. Cell viability was calculated as the ratio between the total number of cells and the number of stained, dead cells.

Freezing of cells

The cells were pelleted by centrifugation at 400xg for 5 min, the supernatant was aspirated, and the cell pellet was re-suspended at room temperature in 1 ml of freezing solution (20% FCS + 10% DMSO in RPMI-1640) in a cryovial (Nunc). The vials were transferred to a styrofoam freezing box and kept overnight at –85°C. On the next day, the vials were transferred to a liquid nitrogen tank.

Thawing of cells

After removal from the liquid nitrogen tank, the frozen cell suspension was quickly thawed in a 37°C water bath (about 1 min). The suspension was diluted in a 15-ml Falcon centrifuge tube with 5 ml of warm growth medium (37°C) and centrifuged at 400xg for 5 min. After the

supernatant was aspirated, the cells were re-suspended in 10ml of medium and transferred in a culture flask.

2.2.2. DNA methods

2.2.2.1. Primer design

Nucleotide sequences of the full genomes of HPV16 (104 genomes), HPV18 (24 genomes), HPV33 (22 genomes), and HPV56 (6 genomes) subtypes were downloaded from the NCBI nucleotide database. Sequences were aligned using the ClustalW software in order to identify regions of the HPV genomes suitable for design of the subtype-specific primers. This analysis revealed that the best region for placing of primers is the end and begin of the *E6* and *E7* open reading frames (ORFs), respectively, because significant, subtype-related diversification of the nucleotide sequences has been noticed in this genomic region. Design of primer pairs was performed using the Clone Manager Professional software (Version 9 for Windows, http://www.scied.com). In order to be able to detect simultaneously HPV subtypes in the clinical samples by a standard PCR followed by an analytical agarose gel electrophoresis, subtype-specific primers were designed to amplify PCR products of different length: 109 bp (HPV16), 84 bp (HPV18), 173 bp (HPV33), and 136 bp (HPV56).

Gene-specific primers were either selected from Primer Bank (http://pga.mgh.harvard.edu/primerbank/index.html) and RTPrimerDB databases (http://medgen.ugent.be/rtprimerdb/index.php) or designed using the Clone Manager Professional software.

Primer name	Sequence (5‘-3‘)
HPV16-E6E7-FW1	GCAGATCATCAAGAACACGTAGAG
HPV16-E6E7-REV1	GTAGAGATCAGTTGTCTCTGGTTG
HPV18-E6E7-FW1	AGCACGACAGGAAAGACTCCAACG
HPV18-E6E7-REV1	GTCTTGCAATGTTGCCTTAGGTCC
HPV33-E6E7-FW1	TGCGGCGTGTTGGAGGTCCCGA
HPV33-E6E7-REV1	TCCAAGCCTTCATCCTCATCTG
HPV56-E6E7-FW1	GTGCTGGAGACAAACATCTAGAG
HPV56-E6E7-REV1	CCAATTGCTCATTGCACTGTAGG

2.2.2.2. Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction

Polymerase chain reaction (PCR) was performed on the 8 (5 samples) pools of 40 cDNA samples from the biopsies and on cDNA from the head and neck cancer cell lines in order to detect the presence of the expressed HPV transcripts. PCR was also performed using the HPV-positive cell lines as a positive control.

DNA amplification via the polymerase chain reaction was performed in the Biometra TProfessional Basic Thermal Cycler. We used the following reagents in a 0.2 ml thin-walled, flat-capped tube (Chemie Brunschwig AG)): 2 µl of template DNA (cDNA), 0.6 µl of forward and reverse primers targeting the region to be amplified, and Go Taq DNA polymerase master mix (Promega; contains GoTaq® DNA Polymerase, 5X Green GoTaq® Reaction Buffer and 5X Colorless GoTaq® Reaction Buffer). The cycling protocol consisted of an initial denaturation at 95°C during 120 seconds and 30 cycles under the following conditions: denaturation at 94°C for 30 sec, annealing of the primers at 55°C for 30 sec, and extension of the primers at 72° C for 30 seconds, and the last step is final extension at 72°C for 400 seconds. After PCR amplification, aliquots of the mixture were loaded onto 1.5% agarose gels (Chemie Brunschwig AG) and electrophoresed as described below. To prepare samples for electrophoresis, 1 µl of SYBR Safe DNA Gel Stain loading buffer was added to every 25 µl of DNA solution. The amplification products were visualized with a mixture of Dye Blue, and their size was estimated using 100bp DNA Ladder (6 µl per lane).

Agarose gel electrophoresis

Agarose gel 1.5% was prepared by heating a quantity of solution (in order to have a thickness of 5-10mm) in a microwave oven. Before pouring the gels, we needed to wait few minutes for the solution to cool to around 50°C. The gel was then prepared by pouring the 50°C warm gel solutions into horizontal trays with inserted combs (BioRad). The gel solidifies at room temperature, generally in 2h. Once the gel becomes solid we removed the combs. The trays were placed in an electrophoresis chamber and covered with electrophoresis buffer (the same buffer used to prepare the agarose solution). Electrophoresis was performed at 90 V during 40 minutes (the time that the marker dye has migrated an appropriate distance that depends on the size of the DNA fragments).

Gene expression analysis of public datasets

For this part of the experiment, we first determine expression of genes from a dataset analysis using Gene Expression Omnibus (GEO) database. GEO is a public collection of all published gene expression datasets. From GEO we selected a dataset including head and neck squamous cell carcinoma (HNSCC) and cervical cancer samples with HPV status (GSE6791 dataset). From this dataset we selected all HPV positive (n=8) and HPV negative (n=13) HNSCC samples for data analysis. Criteria's for groups' comparison were fold changes and t-test. Another dataset that we analyzed was the GSE24089 dataset, which contains expression data from HPV (-) and HPV (+) squamous cell carcinoma cell lines: SCC4 HPV (-), SCC74A HPV (-), SCC47 HPV (+) and CaSki HPV (+). CaSKi is originating from cervix cancer, SCC4, SCC74A and SCC47 are from HNSCC. Criteria's for groups' comparison were also fold changes and t-test. Then some genes up-regulated in the HPV-positive HNSCC samples and HPV-positive cell lines were selected for next qRT-PCR analysis.

qRT-PCR was performed on 4 selected cell lines and 2 tumour samples according to their HPV status, origins and p53 status. qRT-PCR was performed using the SYBR Green assay.

Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
SHISA2	TGTCAGCAGGCCACCCAGATTG	AGCTGTCATGGGCACAGAGTC
PARM1	CAACTCGGTGCTCCCAGTTAC	CGCTTGTTGACGAGAACCCA
TDRD9	AAGAGCTTGATGTGTGTCGCA	TGAGTGCTTTTACCGCTTCCC
TCAM1P	GCAGCTCACCCATCAGAATAG	TGGTCTTCCAGGAACATGTG
MEST	AGGGACTGCGTATCTTCTACC	ACCCGATGAAACCTCAAGGTC
HPRT1 (housekeeping gene)	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT

2.2.3. RNA methods

2.2.3.1. RNA preparation

Total RNA was prepared from cultured cells and tumor tissue samples with the Direct-Zol RNA mini Prep kit (Zymoresearch) according to the manufacturer's instructions. In order to provide appropriate conditions for the binding of the RNA to the silica-gel membrane of the Zymo spin column, we added ethanol to the lysates. The samples were then treated on-column with RNase-free DNase I to remove contaminating genomic DNA. RNAse-free water

was used to elute RNA from the membrane and recover it in a sterile tube. Then we used Nanodrop spechtrophotometry (test type: nucleic acid) to evaluate and measure the concentration and purity of RNA by measuring the absorbance at 260 nm and 280 nm.

2.2.4. Protein methods

2.2.4.1. Protein determination by Bradford Protein Assay

Cell extracts were prepared using KLB-buffer (1ml of complete lysis buffer for a confluent T75 cell culture flask). Cells were scraped in KLB-buffer and incubated 30 minutes on ice, then centrifugated for 10 minutes at 4°C to peelt the debrisand the supernatant was transfered in a 1.5 ml tube. To determine the protein concentration BCA protein assay kit (Pierce) was used (spectrometry analysis at 562 nm wave-length). Then we analyzed the protein concentration using the Biotek Epoch Microplate Spectrophotometer and the Gens1.10 Software.

2.2.4.2. SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Standard SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out in Mini-Protean II dual slab cells (Bio-Rad). For the resolving gel 10% we prepared 10 ml that is sufficient for small (8 cm x 10 cm x 1.5 mm) gel format. For the 5% stacking gel we prepared 3 ml of gel solution. Both gels were prepared with SDS 10 %, TEMED and ammonium persulfate were added just prior to pouring the gel. The separating gel solution was poured into the assembled gel apparatus, overlaid with water and allowed to polymerize. After the resolving gel had polymerized, we put 10 ul of isopropanolol on the gel surface and remove with some blotting paper. Then we poured the staking gel and we inserted the combs and wait for stacking gel polymerization. The protein samples were prepared by heating them 1 minute at 95°C. The gel was run at a constant voltage of 100V for 1h15.

2.2.4.3. Western blotting

Western Blotting was then performed on 4 selected cell lines according to their HPV status, origins and p53 status. HPV positive head and neck cancer cell line was also selected as a positive control comparison. Two proteins were selected for Western Blotting analysis of cell

lysates: rabbit polyclonal anti-SHISA2 antibody (cat. N. PA5-25853) was purchased from Pierce.

The polyacrylamide gel was sandwiched between two layers of filter paper and two sponges in a cassette. Electrophoretic transfer of proteins to nitrocellulose sheets was performed using a Mini Protean II transfer system (Biorad). The nitrocellulose sheet (Protran BA-85, 0.45 µm), filter paper and sponges were first soaked in the transfer buffer. We removed then the gel from the electrophoresis chamber and removed the upper part of the gel (stacking gel), always with liquid to not dry the gel. The gel was equilibrated with the transfer buffer. After assembling of gel/nitrocellulose membrane sandwiched between filter papers and foam pads in a special cassette, the cassette was placed inside of the support frame of the transfer cell such that the nitrocellulose side of the stack was towards the positive pole, because protein are negatively charged and will run from the positive to the negative pole. The cooling box filled with ice was inserted into the apparatus and the appropriate volume of transfer buffer was added. The transfer was performed for 75 minutes at 100V. After completion of the electrotransfer, the power was turned off, the apparatus was disassembled, and the nitrocellulose filter was removed and briefly dried.

To reduce non-specific binding of antibody (that will be used for the detection of the target protein), the nitrocellulose membrane was incubated in a 5% milk-solution by continuously agitating it on a rocking platform for at least 60 min at room temperature (5% non-fat milk in PBS/0.1% Tween 20= blocking buffer). We prepared the primary antibody solution by adding antibody solution in 2.5% milk in PBS/0.1%Tween 20. Then we took the membrane and rinse it with PBS/0.1%Tween 20 and put it in a Falcon tube with the protein face inward the tube and we let it incubate with the primary antibody solution overnight in a cold room on a mixer plate. After 3 washes (15 min each) with PBS/0.1% Tween 20, the blot was incubated for 1 h with the horseradish peroxidase-labelled secondary antibody diluted in the 2.5% milk solution. The appropriate titers of primary and secondary antibodies varied and were determined in preliminary experiments. We used as dilutions 1:100 for primary antibodies and 1:2500 for secondary antibodies. After that, the nitrocellulose sheet was again washed 3 times 8 minutes and incubated with the luminol-based ECL Western blotting detection reagents from Pierce (SuperSignal West Pico). We incubated the membrane 5 minutes, and then we dry the membrane and put it in the RX-detection box. Light generated by the chemiluminescent reaction was detected by exposing X-ray film (Fuji).

3. Results

3.1. Multiplexed RT-PCR method for detection of HPV transcripts in tumor samples and cell lines

RNA from 40 tumor samples originating from oral cavity and oropharynx were collected at the CHUV (Centre Hospitalier Universitaire Vaudois). All surgical tumor samples were collected before any previous adjuvant treatments (radiotherapy or chemotherapy). 40 RNA samples have been converted into cDNA, which were pooled into 8 groups containing 5 cDNA samples. Sample pooling was done to reduce the number of PCR amplifications, because low frequency of HPV-expressing tumor samples was expected. Other RNA samples were derived from head and neck cancer cell lines established in the ORL lab were also converted into cDNA and tested by PCR for the presence of HPV transcripts. To control the specificity of the designed primers we used as a positive control, cDNA from human Hela cells that express HPV 18 transcripts and mouse TC-1 cells that express HPV 16 transcripts. In addition, genomic DNA of HPV 56 subtype (kindly provided by Dr. Roland Sahli, CHUV) was used as control for HPV56-specific primers. Multiplexed PCR amplification was used for detection of expressed HPV transcripts. Four primer pairs specific for HPV 16, HPV18, HPV33, HPV56, designed by Dr. Genrich Tolstonog and described in the material and methods section, were used. Design of primers was done to detect by agarose gel electrophoresis four HPV isotypes based on the difference in size (base pair) of the amplified PCR products: 84 bps for HPV-18, 109 bps for HPV-16, 136 bps for HPV-56 and 173 bps for HPV-33.

We proceeded in two steps. In the first step the eight pools were tested by PCR. We found that one of the samples was HPV-33 positive (Figure 5-A). Therefore, we tested then the individual samples of this pool separately and found that HPV-33 transcipts are detectable in patient number 25, which was a tumor sample collected in the oral cavity (Figure 5-B). This sample will be used for testing of genes from a bioinformatic analysis. In the second step, we tested the cell lines established in the ORL lab also by using PCR and the four primers pairs. As shown in Figure 5-C, no cell lines express HPV subtypes 16, 18, 33 and 56.

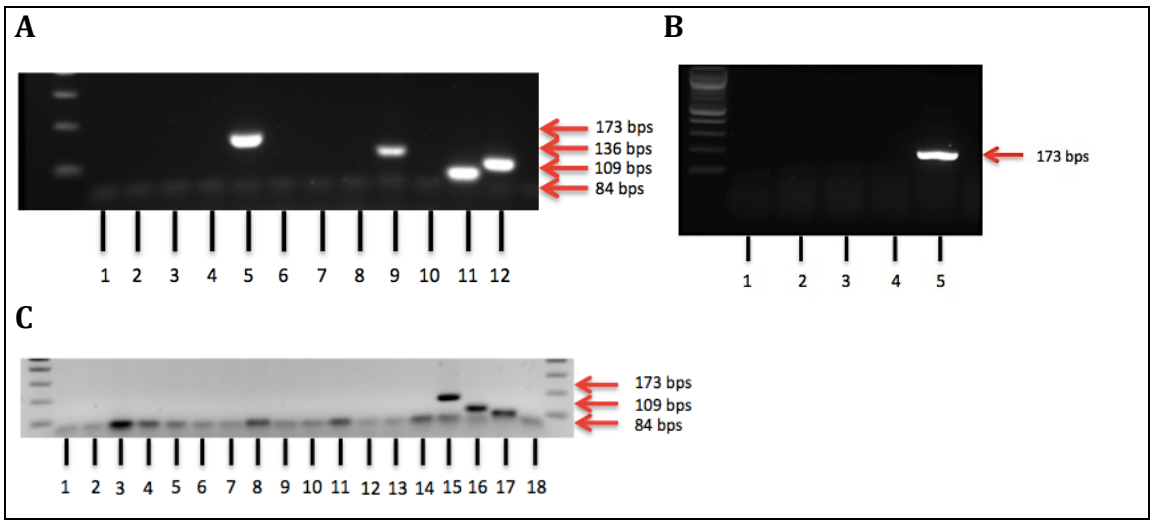

Figure 5: Detection of HPV transcripts in tumor samples biopsies. **A)** PCR performed on the 8 (5 samples) pools of 40 cDNA samples. 3 positive controls (9: HPV56 genomic DNA;11: HPV18 transcripts in Hela cells;12: HPV16 transcripts in TC-1 cells) and one negative control (10: water control) have been used. Positive signal for HPV-33 (173 bps) is in the pool number 5. **B**) PCR performed on the 5 samples from the pool number 5. Positive signal for HPV-33 (173bps) in the biopsies number 5. **C**) PCR performed on cDNA from the head and neck cancer cell lines (1-14). 3 positive controls (15: HPV33 genomic DNA; 16: HPV16 transcripts in TC-1 cells; 17: HPV18 transcripts in Hela cells), 1 negative control (18: water control). No detection of HPV transcripts in the tested cell lines. Band detected in most samples including water control represent primer dimers.

3.2. Bioinformatics gene expression analysis of HNSCC tumor samples and cell lines with known HPV status

In order to identify genes which expression correlates with the HPV status of tumors and is specifically expressed in HPV-positive cancer cell lines, we took advantage of the GEO database that is a public collection of gene expression datasets. From GEO database we selected GSE6791 dataset that includes HNSCC and cervical cancer samples with known HPV status. For bioinformatic study we used gene expression data from 8 HPV-positive and 13 HPV-negative samples in the GSE6791 dataset, and we applied the following criteria for group comparison: fold change (more than 2 fold) and t-test ($p < 0.05$). This analysis has shown that 101 genes represented by 137 Affymetrix probesets are differentially expressed between HPV-positive and HPV-negative HNC samples: 48 genes are up-regulated and 53 genes are down-regulated in the HPV-positive tumor samples (Table 1 and 2). As expected, the list of the top-regulated genes includes *CDKN2A* gene, which encodes p16 protein, a surrogate marker of HPV-positive head and neck tumors. From the list of the up-regulated genes were selected some top-regulated genes for the next experiments.

Table 1: List of genes up-regulated in HPV-positive samples from the GSE6791 dataset

probeid	symbol	NAME	LogFC	adj.pvalue
230493_at	*SHISA2*	shisa homolog 2 (Xenopus laevis)	4.10	6.67E-08
227662_at	*SYNPO2*	synaptopodin 2	4.00	1.70E-07
233320_at	*TCAM1P*	testicular cell adhesion molecule 1 homolog (mouse), pseudogene	3.36	5.74E-15
231164_at	*ABCA17P*	ATP-binding cassette, sub-family A (ABC1), member 17, pseudogene	3.29	4.16E-12
206546_at	*SYCP2*	synaptonemal complex protein 2	3.26	3.94E-08
203881_s_at	*DMD*	dystrophin	3.25	6.04E-07
235343_at	*VASH2*	vasohibin 2	3.24	2.08E-05
229623_at	*TMEM150C*	transmembrane protein 150C	3.20	4.41E-06
203868_s_at	*VCAM1*	vascular cell adhesion molecule 1	3.19	2.17E-07
201650_at	*KRT19*	keratin 19	3.13	1.40E-03
1557570_a_at	*LOC285084*	hypothetical protein LOC285084	3.05	2.09E-09
205890_s_at	*NA*	NA	2.94	1.33E-07
207039_at	***CDKN2A***	**cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)**	**2.94**	**1.71E-04**
229152_at	*C4orf7*	chromosome 4 open reading frame 7	2.86	4.56E-03
228434_at	*BTNL9*	butyrophilin-like 9	2.84	2.94E-09
1556244_s_at	*LOC375196*	hypothetical protein LOC375196	2.70	1.25E-10
204798_at	*MYB*	v-myb myeloblastosis viral oncogene homolog (avian)	2.67	1.94E-06
227702_at	*CYP4X1*	cytochrome P450, family 4, subfamily X, polypeptide 1	2.53	1.89E-04
225809_at	*PARM1*	prostate androgen-regulated mucin-like protein 1	2.52	6.45E-07
1569040_s_at	*FLJ40330*	hypothetical LOC645784	2.50	5.33E-05
202437_s_at	*CYP1B1*	cytochrome P450, family 1, subfamily B, polypeptide 1	2.50	1.65E-03
217028_at	*CXCR4*	chemokine (C-X-C motif) receptor 4	2.41	2.52E-06
1557545_s_at	*RNF165*	ring finger protein 165	2.35	1.42E-06
215028_at	*SEMA6A*	semaphorin 6A	2.35	6.76E-06
205668_at	*LY75*	lymphocyte antigen 75	2.35	5.61E-06
228262_at	*MAP7D2*	MAP7 domain containing 2	2.34	2.36E-08
210538_s_at	*BIRC3*	baculoviral IAP repeat-containing 3	2.34	1.77E-04
204409_s_at	*EIF1AY*	eukaryotic translation initiation factor 1A, Y-linked	2.32	1.24E-02
203685_at	*BCL2*	B-cell CLL/lymphoma 2	2.30	6.46E-06
226811_at	*FAM46C*	family with sequence similarity 46, member C	2.29	4.19E-05
1556300_s_at	*SIM1*	single-minded homolog 1 (Drosophila)	2.29	5.82E-05
205267_at	*POU2AF1*	POU class 2 associating factor 1	2.25	9.79E-04
204533_at	*CXCL10*	chemokine (C-X-C motif) ligand 10	2.23	3.77E-05
222940_at	*SULT1E1*	sulfotransferase family 1E, estrogen-preferring, member 1	2.19	1.38E-03
234994_at	*TMEM200A*	transmembrane protein 200A	2.18	5.61E-04
205159_at	*CSF2RB*	colony stimulating factor 2 receptor, beta, low-affinity	2.16	1.15E-06
213131_at	*OLFM1*	olfactomedin 1	2.15	2.47E-05
205681_at	*BCL2A1*	BCL2-related protein A1	2.14	5.15E-04
209583_s_at	*CD200*	CD200 molecule	2.12	9.86E-06
209708_at	*MOXD1*	monooxygenase, DBH-like 1	2.11	2.89E-05
205242_at	*CXCL13*	chemokine (C-X-C motif) ligand 13	2.10	3.27E-03
233064_at	*ZFR2*	zinc finger RNA binding protein 2	2.10	5.19E-12

228174_at	*SCAI*	suppressor of cancer cell invasion	2.09	7.33E-06
214669_x_at	*IGKC*	immunoglobulin kappa constant	2.08	6.92E-05
218824_at	*PNMAL1*	PNMA-like 1	2.08	6.41E-06
210072_at	*CCL19*	chemokine (C-C motif) ligand 19	2.06	1.24E-04
205328_at	*CLDN10*	claudin 10	2.04	1.17E-05
209289_at	*NFIB*	nuclear factor I/B	2.04	7.22E-04
242539_at	*DIS3L2*	DIS3 mitotic control homolog (S. cerevisiae)-like 2	2.02	1.63E-04

Table 2: List of genes down-regulated in HPV-positive samples from the GSE6791 dataset

probeid	symbol	NAME	LogFC	adj.pvalue
209160_at	*AKR1C3*	aldo-keto reductase family 1, member C3	-4.26	8.98E-06
206561_s_at	*AKR1B10*	aldo-keto reductase family 1, member B10 (aldose reductase)	-3.91	8.45E-09
213506_at	*F2RL1*	coagulation factor II (thrombin) receptor-like 1	-3.64	6.68E-12
205916_at	*S100A7*	S100 calcium binding protein A7	-3.62	1.96E-04
223631_s_at	*C19orf33*	chromosome 19 open reading frame 33	-3.53	4.51E-05
219263_at	*RNF128*	ring finger protein 128	-3.44	8.09E-08
211906_s_at	*SERPINB4*	serpin peptidase inhibitor, clade B (ovalbumin), member 4	-3.25	4.90E-04
206166_s_at	*CLCA2*	chloride channel accessory 2	-3.18	1.28E-05
231771_at	*GJB6*	gap junction protein, beta 6, 30kDa	-3.12	7.51E-04
202345_s_at	*FABP5*	fatty acid binding protein 5 (psoriasis-associated)	-3.04	2.96E-07
230835_at	*KRTDAP*	keratinocyte differentiation-associated protein	-3.00	4.08E-03
206300_s_at	*PTHLH*	parathyroid hormone-like hormone	-2.99	4.29E-07
214612_x_at	*MAGEA6*	melanoma antigen family A, 6	-2.94	2.05E-03
213796_at	*SPRR1A*	small proline-rich protein 1A	-2.88	4.24E-03
209699_x_at	*AKR1C2*	aldo-keto reductase family 1, member C2	-2.88	1.06E-05
209800_at	*KRT16*	keratin 16	-2.67	2.69E-03
222484_s_at	*CXCL14*	chemokine (C-X-C motif) ligand 14	-2.67	1.80E-05
205064_at	*SPRR1B*	small proline-rich protein 1B	-2.66	3.35E-03
209719_x_at	*SERPINB3*	serpin peptidase inhibitor, clade B (ovalbumin), member 3	-2.63	3.56E-04
226926_at	*DMKN*	dermokine	-2.61	2.11E-03
209301_at	*CA2*	carbonic anhydrase II	-2.56	3.52E-05
223278_at	*GJB2*	gap junction protein, beta 2, 26kDa	-2.49	3.14E-04
210503_at	*MAGEA11*	melanoma antigen family A, 11	-2.48	1.57E-03
209942_x_at	*MAGEA3*	melanoma antigen family A, 3	-2.47	2.52E-03
204469_at	*PTPRZ1*	protein tyrosine phosphatase, receptor-type, Z polypeptide 1	-2.47	2.18E-04
214073_at	*CTTN*	cortactin	-2.45	3.90E-05
203980_at	*FABP4*	fatty acid binding protein 4, adipocyte	-2.44	2.47E-03
202990_at	*PYGL*	phosphorylase, glycogen, liver	-2.30	4.23E-08
227736_at	*C10orf99*	chromosome 10 open reading frame 99	-2.28	1.96E-02
203963_at	*CA12*	carbonic anhydrase XII	-2.27	2.29E-06
219795_at	*SLC6A14*	solute carrier family 6 (amino acid transporter), member 14	-2.27	2.88E-03
1562102_at	*AKR1C1*	aldo-keto reductase family 1, member C1	-2.24	5.36E-04
211361_s_at	*SERPINB13*	serpin peptidase inhibitor, clade B (ovalbumin), member 13	-2.23	8.65E-03

210519_s_at	*NQO1*	NAD(P)H dehydrogenase, quinone 1	-2.19	1.74E-05
1558378_a_at	*AHNAK2*	AHNAK nucleoprotein 2	-2.19	7.18E-04
209351_at	*KRT14*	keratin 14	-2.18	1.65E-03
1552487_a_at	*BNC1*	basonuclin 1	-2.17	9.60E-05
205680_at	*MMP10*	matrix metallopeptidase 10 (stromelysin 2)	-2.17	4.08E-04
226863_at	*FAM110C*	family with sequence similarity 110, member C	-2.17	3.18E-05
206354_at	*SLCO1B3*	solute carrier organic anion transporter family, member 1B3	-2.13	3.63E-04
203324_s_at	*CAV2*	caveolin 2	-2.12	4.01E-06
203851_at	*IGFBP6*	insulin-like growth factor binding protein 6	-2.12	5.61E-04
209126_x_at	*KRT6B*	keratin 6B	-2.10	9.42E-03
203535_at	*S100A9*	S100 calcium binding protein A9	-2.09	1.68E-02
232170_at	*S100A7A*	S100 calcium binding protein A7A	-2.09	3.04E-03
204058_at	*ME1*	malic enzyme 1, NADP(+)-dependent, cytosolic	-2.07	1.35E-05
201427_s_at	*SEPP1*	selenoprotein P, plasma, 1	-2.07	1.29E-04
231726_at	*PCDHB14*	protocadherin beta 14	-2.07	8.56E-06
228698_at	*SOX7*	SRY (sex determining region Y)-box 7	-2.06	6.73E-05
227202_at	*CNTN1*	contactin 1	-2.06	2.57E-03
204388_s_at	*MAOA*	monoamine oxidase A	-2.04	8.35E-04
202376_at	*SERPINA3*	serpin peptidase inhibitor, clade A, member 3	-2.04	9.91E-04

The GSE24089 dataset was used to analyze expression data from the HPV-negative and HPV-positive squamous cell carcinoma cell lines: SCC4 HPV (-), SCC74A HPV (-), SCC47 HPV (+) and CaSki HPV (+). SCC4, SCC47 and SCC74A cell lines are derived from head and neck cancer, while CaSKi cell line is derived from cervical cancer. Using the same criteria's for comparison of groups (fold change more than 2 fold) and t-test: $p < 0.05$) we found that 63 genes represented by 74 Affymetrix probesets are differentially expressed between HPV-positive and HPV-negative cell lines: 39 genes are up-regulated and 24 genes are down-regulated in the HPV-positive cell lines (Table 3 and 4). Also this analysis has shown that *CDKN2A* gene is prominently up-regulated in the HPV-positive cell lines. For the next experiments we selected some top up-regulated genes.

Table 3: List of genes up-regulated in HPV-positive cell lines from the GSE24089 dataset

probeid	symbol	NAME	LogFC	adj.pvalue
228285_at	TDRD9	tudor domain containing 9	6.42	0.0016
224918_x_at	MGST1	microsomal glutathione S-transferase 1	6.19	0.0276
205466_s_at	HS3ST1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1	6.06	0.0023
204614_at	SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	5.81	0.0062
209278_s_at	TFPI2	tissue factor pathway inhibitor 2	5.75	0.0022
200953_s_at	CCND2	cyclin D2	5.71	0.0217

207039_at	**CDKN2A**	**cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)**	**5.60**	**0.0026**
215729_s_at	VGLL1	vestigial like 1 (Drosophila)	5.58	0.0041
234973_at	SLC38A5	solute carrier family 38, member 5	5.52	0.0019
219836_at	ZBED2	zinc finger, BED-type containing 2	5.24	0.0062
222242_s_at	KLK5	kallikrein-related peptidase 5	5.19	0.0133
244050_at	PTPLAD2	protein tyrosine phosphatase-like A domain containing 2	5.17	0.0133
223861_at	HORMAD1	HORMA domain containing 1	5.14	0.0062
227850_x_at	CDC42EP5	CDC42 effector protein (Rho GTPase binding) 5	5.10	0.0031
202016_at	MEST	mesoderm specific transcript homolog (mouse)	5.08	0.0042
219181_at	LIPG	lipase, endothelial	5.04	0.0146
1552797_s_at	PROM2	prominin 2	4.97	0.0276
213113_s_at	SLC43A3	solute carrier family 43, member 3	4.85	0.0486
210118_s_at	IL1A	interleukin 1, alpha	4.78	0.0053
206595_at	CST6	cystatin E/M	4.76	0.0133
226847_at	FST	follistatin	4.73	0.0276
220027_s_at	RASIP1	Ras interacting protein 1	4.63	0.0117
223784_at	TMEM27	transmembrane protein 27	4.62	0.0241
204465_s_at	INA	internexin neuronal intermediate filament protein, alpha	4.61	0.0166
205206_at	KAL1	Kallmann syndrome 1 sequence	4.60	0.0090
221648_s_at	NA	NA	4.52	0.0133
207517_at	LAMC2	laminin, gamma 2	4.34	0.0133
238439_at	ANKRD22	ankyrin repeat domain 22	4.28	0.0133
233320_at	TCAM1P	testicular cell adhesion molecule 1 homolog (mouse), pseudogene	4.15	0.0241
236918_s_at	LRRC34	leucine rich repeat containing 34	4.01	0.0486
210550_s_at	RASGRF1	Ras protein-specific guanine nucleotide-releasing factor 1	3.98	0.0449
208291_s_at	TH	tyrosine hydroxylase	3.92	0.0242
229332_at	HPDL	4-hydroxyphenylpyruvate dioxygenase-like	3.88	0.0241
1559607_s_at	GBP6	guanylate binding protein family, member 6	3.87	0.0342
226145_s_at	FRAS1	Fraser syndrome 1	3.83	0.0347
226132_s_at	MANEAL	mannosidase, endo-alpha-like	3.82	0.0276
210038_at	PRKCQ	protein kinase C, theta	3.78	0.0276
1556244_s_at	LOC375196	hypothetical protein LOC375196	3.77	0.0418
211712_s_at	ANXA9	annexin A9	3.53	0.0404

Table 4: List of genes down-regulated in HPV-positive cell lines from the GSE24089 dataset

probeid	symbol	NAME	LogFC	adj.pvalue
204597_x_at	STC1	stanniocalcin 1	-7.18	0.0003
201744_s_at	LUM	lumican	-5.96	0.0016
204337_at	RGS4	regulator of G-protein signaling 4	-5.72	0.0241
219288_at	C3orf14	chromosome 3 open reading frame 14	-5.66	0.0019
221730_at	COL5A2	collagen, type V, alpha 2	-5.58	0.0062
213425_at	WNT5A	wingless-type MMTV integration site family, member 5A	-5.57	0.0296
212094_at	PEG10	paternally expressed 10	-5.52	0.0019
205767_at	EREG	epiregulin	-5.42	0.0241

206172_at	IL13RA2	interleukin 13 receptor, alpha 2	-5.41	0.0313
218573_at	MAGEH1	melanoma antigen family H, 1	-4.91	0.0041
226875_at	DOCK11	dedicator of cytokinesis 11	-4.62	0.0090
230563_at	RASGEF1A	RasGEF domain family, member 1A	-4.55	0.0174
226051_at	SELM	selenoprotein M	-4.50	0.0296
219985_at	HS3ST3A1	heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1	-4.45	0.0174
209230_s_at	NUPR1	nuclear protein, transcriptional regulator, 1	-4.26	0.0276
211653_x_at	AKR1C2	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2	-4.19	0.0276
205741_s_at	DTNA	dystrobrevin, alpha	-3.99	0.0312
218638_s_at	SPON2	spondin 2, extracellular matrix protein	-3.95	0.0497
212154_at	SDC2	syndecan 2	-3.92	0.0217
204151_x_at	AKR1C1	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1	-3.85	0.0486
231725_at	PCDHB2	protocadherin beta 2	-3.70	0.0347
232195_at	GPR158	G protein-coupled receptor 158	-3.70	0.0325
209160_at	AKR1C3	aldo-keto reductase family 1, member C3	-3.67	0.0486
210102_at	VWA5A	von Willebrand factor A domain containing 5A	-3.65	0.0347

At this point the selected 5 genes for further validation by qRT-PCR: *SHISA2* (up-regulated in HPV-positive samples from the GSE6791 dataset), *PARM1* (up-regulated in HPV-positive samples from the GSE6791 dataset), *TDRD9* (up-regulated in HPV-positive samples from the GSE24089 dataset), *TCAMP1* (up-regulated in HPV-positive samples from the GSE6791 dataset) and MEST (up-regulated in HPV-positive samples from the GSE24089 dataset).

Shisa family member 2 (*SHISA2*) gene encodes a protein that plays an essential role in the maturation of mesoderm cells by individual attenuation of both FGF and WNT signaling. Prostate Androgen-Regulated Mucin-Like Protein 1 (*PARM1*) is a protein-coding gene which function has an implication in the regulation of telomerase activity, that enabling certain prostate epithelial cells to resist apoptosis. PARM1 is also known to be associated with certain diseases including epispadias. Tudor Domain Containing 9 (*TDRD9*) is a protein-coding gene. It has a putative ATP-binding RNA helicase activity that plays a role during spermatogenesis and it is essential for the germ line integrity. Testicular cell adhesion molecule 1 (*TCAM1P*) is a non-coding transcript. In mouse, *TCAM1* gene is known to encode a testis-specific cell adhesion protein that may play a role in germ cell-Sertoli cell interactions. It regulates gene expression and it is active in stem cells. Mesoderm specific transcript (*MEST*) is mono-allelically expressed and epigenetically regulated non-coding gene. *MEST* function is implicated into genetic imprinting to regulate genes. It is known to

have a role in gene regulation in fetal tissues and in imprinting in lymphocytes. The loss of imprinting has been linked to different cancer types.

We tested expression of the selected genes by using real-time qRT-PCR technique in 5 different cell lines that we selected according to their HPV status (positive or negative), localization (oral cavity or oropharynx) and p53 status (wild-type or mutated). In addition we tested the selected genes in 2 tumor samples selected according to their HPV status (positive or negative). Another cell line, MDAmb231, was selected from a CCLE database that contains data for gene expression across 1000 cell lines by criteria of high expression level of *SHISA2* gene. MDAmb231 is breast cancer cell line derived from metastatic site (pleural effusion) and is featured by a mesenchymal phenotype (high expression of EMT-related proteins, high invasion and metastasis *in vivo*). MDAmb231 was used as a positive control for *SHISA2* gene expression in our experiment. All these samples are presented in material and methods section.

The results of our experiment are presented in Figure 6. The up-regulation of *SHISA2* is clearly shown in the HPV positive cell line (SCC-090) and in the HPV positive clinical sample (25). That makes *SHISA2* gene as a potential marker of HPV-expressing tumors. *SHISA2* is also up-regulated as expected in MDAmb231 cell line, indicating that this positive control works well.

PARM1 gene is up regulated in several cell lines and in the HPV negative clinical sample. It does not seem to be linked to HPV specifically, and we conclude that *PARM1* gene is not suitable as a marker candidate.

TDRD9 gene is exclusively expressed in HPV positive cell line (SCC-090) and not in the HPV positive clinical sample. One explanation for this result can be the fact that tumor is a complex tissue containing heterogeneous population of cells with different proliferation potential. In contrast, the established cell lines are more homogeneous in respect to their proliferative activity. Therefore, it is tempting to speculate that *TDRD9* gene expression is associated with the proliferative fraction of HPV-positive tumor cells.

TCAM1P gene is clearly expressed in HPV positive cell line SCC-090 and in LAU-2092 cell line. Interestingly, LAU-2092 cell line is originating from oropharynx, where the most HPV-induced tumors originate. One explanation for this observation is that the expression of some genes depends on the cell localization, so maybe *TCAM1P* has not the same expression

in oral cavity as in oropharynx. Therefore, we conclude that *TCAM1P* gene seems to be linked to oropharynx than to HPV.

MEST gene is up-regulated in many cell lines without any specificity to HPV, therefore it is not considered as a potential marker.

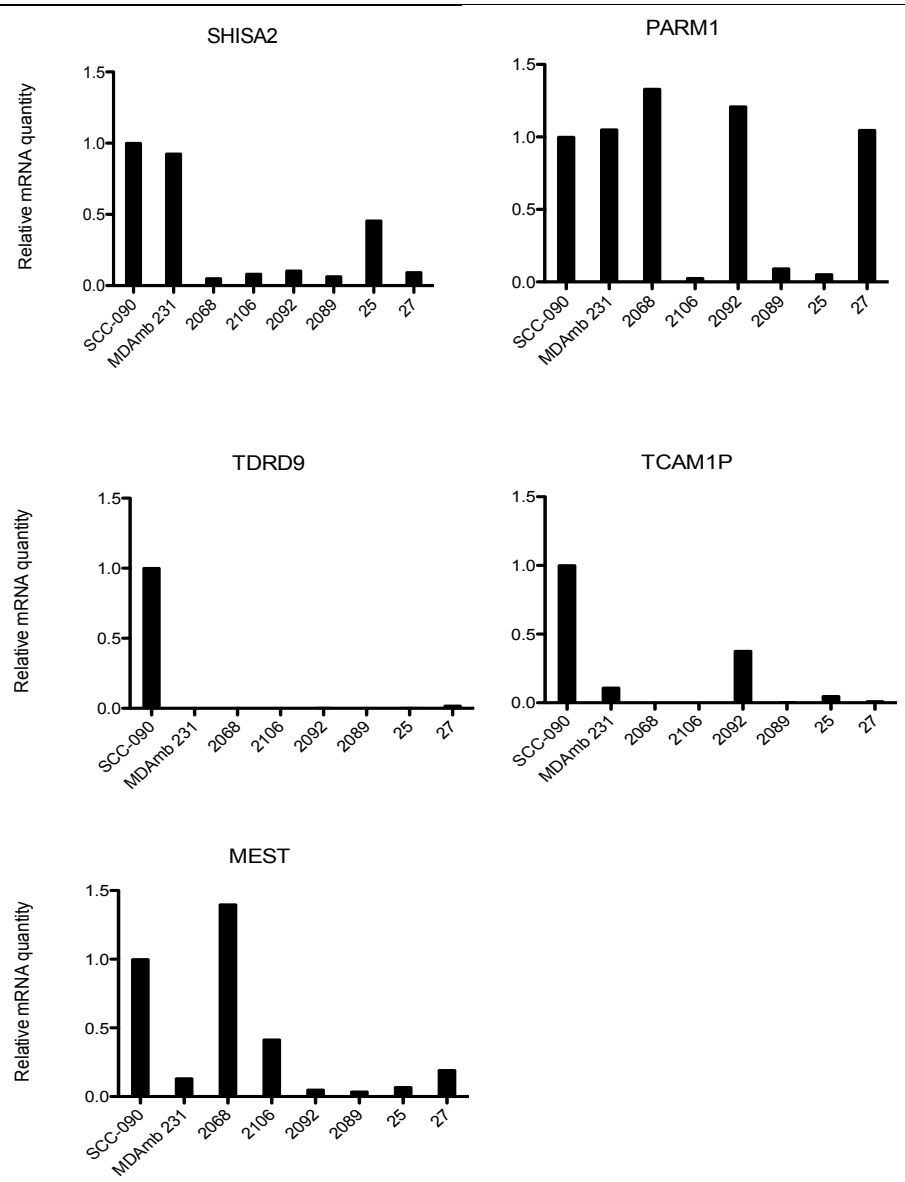

Figure 6: qRT-PCR analysis of gene expression.
Genes selected by bioinformatic analysis were tested using Real-time qRT-PCR with total RNA extracted from 6 head and neck cancer cell lines, MDAmb231 cell line and 2 tumor samples differing in HPV status.
SHISA2 gene is clearly up-regulated in HPV-positive cell line and HPV-positive tumor sample. *PARM1* gene is up-regulated in many cell lines and tumor samples. *TDRD9* gene is exclusively expressed in HPV-positive cell line. *TCAM1P* gene is up-regulated in HPV-positive cell line. *MEST* gene is up-regulated in many cell lines.

3.3. Validation of SHISA2 protein expression by western blotting

According to previous experiments, we selected *SHISA2* gene that showed to be linked with HPV. Applying Western blotting we wanted to test if *SHISA2* gene is expressed as protein in the tested cell lines and tumor samples. We tested a commercially available SHISA2 antibody, which is distributed by multiple companies. According to the manufacturer, this antibody has been raised against synthetic peptide between 259-288 amino acids from the C-terminal region of human SHISA2 protein. The theoretical molecular weight of SHISA2 precursor protein is 31.4 kDa (295 aa) and molecular weight of SHISA2 protein after cleavage of the N-terminal signaling peptide is 28.3 kDa (265 aa), however, as shown in Figure 7 the SHISA2 antibody detects multiple proteins of higher molecular weight and no bands correspond to the predicted protein size. This result indicates that the selected antibody is non-specific and another antibody with a validated specificity should be applied to test SHISA2 protein expression in HPV-positive and –negative head and neck cancer samples.

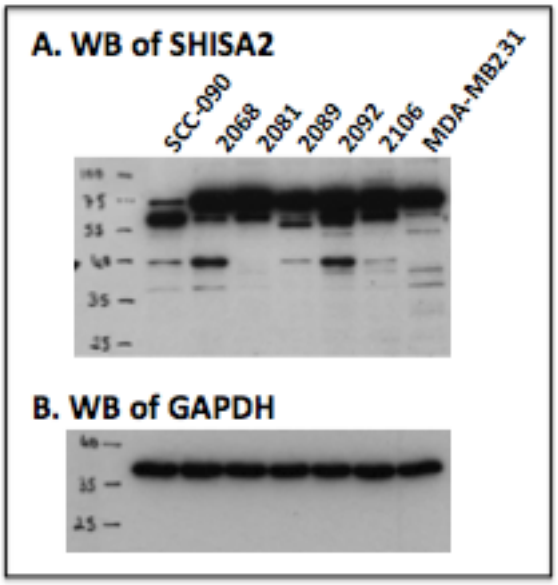

Figure 7: Western blotting analysis of SHISA2 protein expression.

(A) Equal protein amounts (15 µg) of seven samples have been electrophoresed in 10% polyacrylamide gel and after transfer into nitrocellulose membrane probed with SHISA2 antibody (PA5-25853; 1:2500).

(B) GAPDH antibody (MAB374; 1:5000) served as control for equal loading.

4. Conclusion and discussion

Multiplexed RT-PCR-based analysis of HPV transcripts in tumor samples has shown that HPV 33 expression in only one tumor sample. It is surprising because it is know that the proportion of oral cancers linked to HPV is increasing, particularly in the USA. One speculation for these unexpected results is that there are cultural differences between Europe and USA, particularly in sexual education, law of health insurance and sexual

behaviors. We need also to consider that only four HPV isotypes are covered by our multiplexed primers and other HPV isotypes related to oral cancer should be included in this study. Alternative explanation for low frequency of the HPV-positive HNSCC cases is the nature of the method used to detect HPV in the clinical samples. Our method is based on the detection of HPV transcripts, and therefore the cases of transcriptionally silent HPV genomes integrated into the host genome during the initiating phase of HPV-induced oral tumorigenesis are not detected by this technique. For systematic analysis of HPV occurrence in the clinical samples, different methods of HPV detection should be combined. Detection of HPV genome integration by the diverse techniques available today (e.g. by HPV subtype-specific PCR from the genomic DNA or *in situ* hybridization on paraffin sections with the specific probes) needs to be used in combination with our RNA-based analysis to reliably detect the cases of HPV-induced and HPV-driven oral cancer.

In the part of validation of genes selected from bioinformatics analysis, we found that at least two genes are promising candidates for future work on the HPV-induced HNSCC: *SHISA2* and *TDRD9*. *SHISA2* gene has shown a good correlation with HPV and was up-regulated in HPV positive cell line and in tumor samples HPV positive. The fact that it is also up-regulated in mesenchymal breast cancer cell line makes us speculate that SHISA2 has probably a role in the regulation of epithelial-mesenchymal transition (EMT). EMT means that cells lose their polarity and the cell-cell adhesion. Acquisition of mesenchymal phenotype gives the cells gain of mobility enabling migration and invasion. This cell state is often found in the metastatic cells. This could be a possible clue to target metastasis in HPV-positive HNSCC.

TDRD9 gene is clearly expressed in HPV-positive cell line but not in HPV-positive tumor sample. One explanation is that maybe *TDRD9* is specifically expressed by cycling tumor cells, which fraction could be much higher in the cell culture than in the tumors. Alternatively, TDRD9 may be specifically expressed by cells expressing stem cell phenotype, which frequency could be also higher in the cell culture than in the tumors. Other HPV-positive cell lines need to be tested for expression of *TDRD9* gene and potential role of TDRD9 protein in cell proliferation as well as for proposed function of TDRD9 in the cancer stem properties of HPV-induced NHSCC. This still has to be answered and other studies are needed.

PARM1 and *TCAM1P* genes very strongly expressed in 2092 cell line, which was derived from oropharynx tumor. These genes don't seem to be linked to HPV but more to the localization of the cancer, which is in this case oropharynx.

At the end of this study, we conclude that HPV–positive HNSCC warrants further studies. The results of our bioinformatic study are promising, as there are other not yet validated genes that are putatively linked to HPV-induced tumorigenesis and their significance need to be tested in functional assays employing HPV-expressing HNSCC cell lines. It is to expect that further validation of differentially expressed genes will lead to identification of potential markers of HPV-associated oral cancer. The functional assays can potentially lead to design of new treatment strategies of HPV-induced HNSCC.

The initial aim of this master thesis was to detect HPV-expressing samples, and then to analyze genes up-regulated in HPV-positive tumors in order to better understand mechanisms of HPV-induced oral cancer. Our study is supporting a characteristic and unique transcriptional profile in HPV-induced HNSCC and the importance to consider and treat this disease as an independent cancer entity. Further studies need to be performed to better understand HPV-induced carcinogenesis and clinical behavior of HPV-positive cancers in order to improve disease prevention, patient care, and to develop new therapeutic strategies for HNSCC patients.

Bibliography

1. Leemans CR, Braakhuis BJM, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer. janv 2011;11(1):9-22.

2. Pai SI, Westra WH. Molecular Pathology of Head and Neck Cancer: Implications for Diagnosis, Prognosis, and Treatment. Annu Rev Pathol Mech Dis. févr 2009;4(1):49-70.

3. Snow AN, Laudadio J. Human papillomavirus detection in head and neck squamous cell carcinomas. Adv Anat Pathol. nov 2010;17(6):394-403.

4. Evans M, Powell NG. The Changing Aetiology of Head and Neck Cancer: the Role of Human Papillomavirus. Clin Oncol. sept 2010;22(7):538-46.

5. Ang KK, Sturgis EM. Human Papillomavirus as a Marker of the Natural History and Response to Therapy of Head and Neck Squamous Cell Carcinoma. Semin Radiat Oncol.avr 2012;22(2):128-42.

6. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer J Int Cancer. 15 juin 2006;118(12):3030-44.

7. Ma J, Liu Y, Huang X-L, Zhang Z-Y, Myers JN, Neskey DM, et al. Induction chemotherapy decreases the rate of distant metastasis in patients with head and neck squamous cell carcinoma but does not improve survival or locoregional control: a meta-analysis. Oral Oncol. nov 2012;48(11):1076-84.

8. Chen R, Aaltonen L-M, Vaheri A. Human papillomavirus type 16 in head and neck carcinogenesis. Rev Med Virol. nov 2005;15(6):351-63.

9. Chung CH, Gillison ML. Human Papillomavirus in Head and Neck Cancer: Its Role in Pathogenesis and Clinical Implications. Clin Cancer Res. 15 nov 2009;15(22):6758-62.

10. Campisi G, Panzarella V, Giuliani M, Lajolo C, Di Fede O, Falaschini S, et al. Human papillomavirus: its identity and controversial role in oral oncogenesis, premalignant and malignant lesions (review). Int J Oncol. avr 2007;30(4):813-23.

11. Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer. août 2010;10(8):550-60.

12. Miller DL, Puricelli MD, Stack MS. Virology and molecular pathogenesis of HPV (human papillomavirus)associated oropharyngeal squamous cell carcinoma. Biochem J. 15 avr 2012;443(2):339-53.

13. Howie HL, Katzenellenbogen RA, Galloway DA. Papillomavirus E6 proteins. Virology. 20 févr 2009;384(2):324-34.

14. McLaughlin-Drubin ME, Münger K. The human papillomavirus E7 oncoprotein. Virology. févr 2009;384(2):335-44.

15. Van Monsjou H s., van Velthuysen M l. f., van den Brekel M w. m., Jordanova E s., Melief C j. m., Balm A j. m. Human papillomavirus status in young patients with head and neck squamous cell carcinoma. Int J Cancer. 15 avr 2012;130(8):1806-12.

16. Rautava J, Syrjänen S. Biology of human papillomavirus infections in head and neck carcinogenesis. Head Neck Pathol. juill 2012;6 Suppl 1:S3-15.

17. Gillison ML, Zhang Q, Jordan R, Xiao W, Westra WH, Trotti A, et al. Tobacco Smoking and Increased Risk of Death and Progression for Patients With p16-Positive and p16-Negative Oropharyngeal Cancer. J Clin Oncol. 10 juin 2012;30(17):2102-11.

18. Ben Aissa A, Mach N. [HPV-related cancer: should young men be vaccinated?]. Rev Médicale Suisse. 23 mai 2012;8(342):1087-90.

19. Olthof NC, Straetmans JMJAA, Snoeck R, Ramaekers FCS, Kremer B, Speel E-JM. Next-generation treatment strategies for human papillomavirus-related head and neck squamous cell carcinoma: where do we go?: Therapies for HPV-related head and neck cancers. Rev Med Virol. mars 2012;22(2):88-105.

20. Mehra R, Ang KK, Burtness B. Management of human papillomavirus-positive and human papillomavirus-negative head and neck cancer. Semin Radiat Oncol. juill 2012;22(3):194-7.

21. Schlecht N, Burk R, Adrien L, Dunne A, Kawachi N, Sarta C, et al. Gene expression profiles in HPV-infected head and neck cancer. J Pathol. nov 2007;213(3):283-93.

22. Pyeon D, Newton MA, Lambert PF, den Boon JA, Sengupta S, Marsit CJ, et al. Fundamental Differences in Cell Cycle Deregulation in Human Papillomavirus-Positive and Human Papillomavirus-Negative Head/Neck and Cervical Cancers. Cancer Res. 15 mai 2007;67(10):4605-19.

Acknowledgments

I would like to thanks…

…Professor Simon for allowing me to gain laboratory experiences and to support me for this project, but also for his encouragement and passion for teaching.

…Dr. Genrich Tolstonog for his support and great help during this project, and also for his patience and unfailing optimism.

…Dr Agnes Hiou for her patience, kindness and teaching.

…The whole team of the laboratory: Dr Jean-Paul Rivals, Dr Jana Sponarova, Dr Vinko Misetic and Philippe Piccardi for helping me not to get lost in the depths of the laboratory.

…All friends and my sister Mélisande that support me.